Niveau(x) d'enseignement: 4 Numéro EDUQ : 8270242
Centre d'intérêt: 08

Bédard, Roger.

La normalisation d'un test par échantillonnage combiné des items et des sujets : effets sur l'estimation des quatre premiers moments et sur la performance à certains items / Roger Bédard. -- [S.l.] : R. Bédard, 1979.
120 p. ; 28 cm.

En tête du titre: Faculté des sciences de l'éducation. Résumé de la thèse (Ph. D.) - Université Laval. Bibliographie: p. [113]-120.

Cote Dewey: 153.93/B399n Cote LC: BF/431

Descripteur(s): *Test d'intelligence *Echantillonnage *Normalisation *Analyse statistique *Secondaire (Niveau)

Identificateur(s) de nom: *Multiple matrix sampling *Echantillonnage combiné des items et des sujets *ECIS

Recherche sur le **Multiple matrix sampling**, une technique de cueillette et d'analyse de données sur échantillons traduite ici par **Echantillonnage combiné des items et des sujets** (ECIS). La recherche a pour but, lorsqu'une telle technique de collecte des données est utilisée dans un contexte de normalisation d'un test d'intelligence: 1. d'étudier les estimations ainsi obtenues des quatre premiers moments; 2. de comparer le rendement des répondants selon la position des items dans le test. Elle vise à une meilleure compréhension de l'effet du contexte ECIS sur l'estimation des paramètres et la performance des répondants.

L'auteur a conçu un test de 72 items à choix multiple extraits de deux versions d'un test d'intelligence déjà validé pour une population du même âge que celle étudiée. Les élèves de Secondaire I de quatre commissions scolaires régionales du Québec formèrent la population de référence. Le test fut normalisé avec les trois plans ECIS sur trois échantillons représentatifs distincts. Concurremment, un autre échantillon de 171 répondants représentatifs passa le test intégral; c'est avec cet échantillon que furent simulés les plans ECIS. L'expérimentation s'est déroulée en mai-juin 1978.

Cette recherche tend à confirmer les études antérieures selon lesquelles l'estimation de la variance par la technique ECIS est suffisamment fiable. Par ailleurs, le réarrangement des items et la variation de la durée de l'épreuve ne seraient pas des facteurs influençant sensiblement la performance des répondants.

Type de document: r
Auteur de l'analyse: jr

E D U Q 8270242

FACULTE DES SCIENCES DE L'EDUCATION

THESE

PRESENTEE

A L'ECOLE DES GRADUES

DE L'UNIVERSITE LAVAL

POUR L'OBTENTION

DU GRADE DE PHILOSOPHIAE DOCTOR

PAR

ROGER BEDARD

MAITRE EN SCIENCES DE L'EDUCATION

DE L'UNIVERSITE LAVAL

LA NORMALISATION D'UN TEST PAR ECHANTILLONNAGE COMBINE DES ITEMS ET DES SUJETS: EFFETS SUR L'ESTIMATION DES QUATRE PREMIERS MOMENTS ET SUR LA PERFORMANCE A CERTAINS ITEMS

JUIN 1979

E D U Q 8270242

RESUME

FACULTE DES SCIENCES DE L'EDUCATION

THESE
PRESENTEE
A L'ECOLE DES GRADUES
DE L'UNIVERSITE LAVAL
POUR L'OBTENTION
DU GRADE DE PHILOSOPHIAE DOCTOR
PAR

ROGER BEDARD
MAITRE EN SCIENCES DE L'EDUCATION
DE L'UNIVERSITE LAVAL

LA NORMALISATION D'UN TEST PAR ECHANTILLONNAGE COMBINE DES ITEMS ET DES SUJETS: EFFETS SUR L'ESTIMATION DES QUATRE PREMIERS MOMENTS ET SUR LA PERFORMANCE A CERTAINS ITEMS

JUIN 1979

Normaliser un test pour une population consiste à déterminer les paramètres de la performance collective de tous les membres de cette population à ce test. En recueillant des données fragmentaires par administration de sous-tests issus du test intégral à autant de sous-échantillons de la population, la technique d'échantillonnage combiné des items et des sujets (ECIS) estime les paramètres de la population parente. Les occasions où cette technique a vraiment servi à la collecte des données sont rares et l'hypothèse d'un effet dû au contexte ECIS d'application sur le rendement des répondants inquiète toujours les utilisateurs. En effet, lorsque les items du test intégral sont distribués dans les sous-tests, plusieurs caractéristiques du contexte d'application de l'épreuve subissent des changements, notamment la réduction de la durée de l'épreuve et le réarrangement des items. Jusqu'ici, les recherches sont peu concluantes.

En étudiant cet effet du contexte ECIS, nous voulons répondre, à la lumière de résultats expérimentaux, aux questions suivantes: 1°) *Quelle influence les modifications ECIS au contexte ont-elles sur les estimations de la moyenne, de la variance et des moments d'ordre 3 et 4?* 2°) *Les modifications ECIS au contexte influencent-elles différemment le rendement des répondants aux items selon que ces derniers se trouvent au début ou à la fin du test intégral?* Chaque question de base est traitée par un plan expérimental distinct, quoique une seule expérimentation permette de recueillir les données propres à chacun.

Tout d'abord, nous avons construit un test de 72 items à choix multiple extraits de deux versions d'un test d'intelligence déjà validé pour une population du même âge que la nôtre. Les indices de facilité des items étant connus, au moment de leur sélection nous avons rendu équivalents les premier et dernier quarts du test par appariement des items. Les élèves de secondaire I de quatre commissions scolaires du Québec formèrent la population de référence. Le test fut normalisé avec les trois plans ECIS, soit (12,6), (4,18) et (2,36), sur trois échantillons représentatifs distincts. Concurremment, un autre échantillon représentatif passa le test intégral; c'est avec cet échantillon que furent simulés les plans ECIS.

En recueillant les données avec les plans ECIS (12,6) et (4,18), dans les deux cas les estimateurs ECIS réel et ECIS simulé de la moyenne s'écartaient significativement l'un de l'autre. Avec le plan ECIS (2,36), les deux estimateurs semblaient estimer la moyenne d'une même population. Pour estimer tant la variance que le moment d'ordre 3, quand elle est appliquée véritablement à la collecte des données, la technique ECIS semble fiable; du moins, nos résultats n'infirment pas cet avancé. Préférant attribuer à une erreur d'échantillonnage l'écart significatif constaté pour le plan (2,36), nous pensons que le moment d'ordre 4 peut aussi être estimé avec sûreté par l'utilisation d'ECIS.

Quant à la variation des indices de facilité, puisqu'une seule différence significative a été constatée dans le cas du plan ECIS (12,6) réel, il apparaît préférable de l'attribuer à des sources expérimentales d'erreur. On peut donc raisonnablement supposer que l'application de la technique ECIS n'a pas d'effet localisé sur la performance de répondants dans des conditions analogues à celles de notre expérience.

Cette recherche tend à confirmer les études antérieures selon lesquelles l'estimation de la variance par la technique ECIS est suffisamment fiable. Nous retenons aussi l'hypothèse que l'on peut estimer sans biais significatif le moment d'ordre 3; certes, l'absence d'études pour la corroborer nous invite à la prudence. Ceci soulève la question du plan ECIS idéal pour éviter de biaiser l'estimation de la moyenne; tenant compte des suggestions de Shoemaker (1973, pages 21 à 23) pour un test de 72 items comme le nôtre, des plans ECIS (6,12) ou (8,9) sans répétition d'items dans un même test ou, encore mieux, des plans (8,18) ou (9,16) à items répétés dans plusieurs tests auraient de bonnes chances d'éviter l'écueil de l'effet de contexte. Par ailleurs, le réarrangement des items et la variation de la durée de l'épreuve ne seraient pas des facteurs influençant sensiblement la performance des répondants.

Vidya Bhushan
Directeur

Roger Bédard
Etudiant

FACULTE DES SCIENCES DE L'EDUCATION

THESE

PRESENTEE

A L'ECOLE DES GRADUES

DE L'UNIVERSITE LAVAL

POUR L'OBTENTION

DU GRADE DE PHILOSOPHIAE DOCTOR

PAR

ROGER BEDARD

MAITRE EN SCIENCES DE L'EDUCATION

DE L'UNIVERSITE LAVAL

LA NORMALISATION D'UN TEST PAR ECHANTILLONNAGE COMBINE DES ITEMS ET DES SUJETS: EFFETS SUR L'ESTIMATION DES QUATRE PREMIERS MOMENTS ET SUR LA PERFORMANCE A CERTAINS ITEMS

JUIN 1979

REMERCIEMENTS

Le présent travail s'est étendu sur une durée de sept ans. Il a fallu, au cours d'une si longue période, solliciter l'appui de nombreuses personnes et à des titres divers. La liste de tous ces collaborateurs et collaboratrices occuperait trop de place. Aussi, je me contenterai de les remercier tous et toutes du fond du coeur, tout en m'engageant à leur réitérer ma gratitude à la première occasion.

J'offre à mon épouse et à mes enfants ce fruit de mes efforts en reconnaissance de la foi qu'ils m'ont témoignée.

De même, je désire remercier tout particulièrement mon directeur de thèse, M. Vidya Bhushan, qui m'a accordé sa confiance et qui a su motiver ma réflexion constamment au cours de ce travail.

MM. Gérard Scallon et Marc-André Nadeau, de qui j'ai reçu les précieux conseils, voudront bien également trouver ici l'expression de mes sincères remerciements.

Roger Bédard

TABLE DES MATIERES

Page

REMERCIEMENTS II

TABLE DES MATIERES III

LISTE DES TABLEAUX V

INTRODUCTION

La technique 1
Le contexte d'utilisation de la technique 3
Le problème 4
Le but de la recherche 5
Les limites de la recherche 6

Chapitre

I LES ECRITS

Les recherches *ex post facto* 9
Les recherches *a priori* 21
Les recherches sur l'effet de contexte 26
Autres recherches 35
Conclusion 38

II LE PLAN DE L'EXPERIENCE

Les plans expérimentaux

Le premier plan expérimental 40
Le deuxième plan expérimental 44

III L'EXPERIENCE

Le test 49
La population 50
Les échantillons de sujets 51
Les plans ECIS 52
La simulation des plans ECIS 53
L'expérimentation 56

IV LES RESULTATS

Le premier plan expérimental 60
Le deuxième plan expérimental 71

Page

CONCLUSION

Résumé 81
Considérations sur la recherche 84
Recommandations au sujet de la technique ECIS 87

APPENDICE A : Le test et les sous-tests 88

APPENDICE B : Les instructions au professeur-collaborateur . . . 102

APPENDICE C : Tableau des indices de facilité des 72 items pour chaque échantillon 107

BIBLIOGRAPHIE 112

LISTE DES TABLEAUX

TABLEAU		Page
1	Sommaire des résultats des recherches *ex post facto* comparant la technique ECIS à la technique EUS pour estimer les paramètres.....	19
2	Répartition hypothétique des sujets par commission scolaire et par plan ECIS.....	51
3	Nombre de groupes-classes, nombre total de sujets et nombre d'observations par plan ECIS et au test intégral, tirés de l'expérimentation.....	52
4	Subdivision du test intégral en sous-tests pour chacun des plans ECIS.....	54
5	Le nombre d'items, le nombre de répondants et le nombre d'observations pour chaque plan ECIS et chaque sous-test, application réelle ou simulée..	57
6	Fréquence des groupes-classes par période d'administration du test et des sous-tests.....	58
7	Les estimations de paramètres et leurs erreurs-types en fonction des plans ECIS (réel ou simulé).	61
8	Analyse de la variance entre les moyennes de l'échantillon de sujets au plan ECIS (12,6) réel et de l'échantillon de sujets au plan ECIS (12,6) simulé.....	62
9	Analyse de la variance entre les moyennes de l'échantillon de sujets au plan ECIS (4,18) réel et de l'échantillon de sujets au plan ECIS (4,18) simulé.....	64
10	Analyse de la variance entre les moyennes de l'échantillon de sujets au plan ECIS (2,36) réel et de l'échantillon de sujets au plan ECIS (2,36) simulé.....	65
11	Les coefficients d'asymétrie A_3 des distributions des échantillons aux plans ECIS réel ou simulé....	68
12	Les coefficients d'aplatissement A_4 des distributions des échantillons aux plans ECIS réel ou simulé.....	69

TABLEAU		Page
13	Analyse de la variance des indices de facilité des items du premier et du dernier quarts du test intégral au plan ECIS (12,6) réel	71
14	Analyse de la variance des indices de facilité des items du premier et du dernier quarts du test intégral au plan ECIS (4,18) réel.....	72
15	Analyse de la variance des indices de facilité des items du premier et du dernier quarts du test intégral au plan ECIS (2,36) réel.....	72
16	Analyse de la variance des indices de facilité des items du premier et du dernier quarts du test intégral pour les répondants aux 72 items.....	73
17	Analyse de la variance des indices de facilité des items 1 à 18 répondus dans les sous-tests de six items et dans le test intégral.....	74
18	Analyse de la variance des indices de facilité des items 55 à 72 répondus dans les sous-tests de 18 items et dans le test intégral.....	75
19	Analyse de la variance des indices de facilité des items 1 à 19 répondus dans les sous-tests de 18 items et dans le test intégral.....	76
20	Analyse de la variance des indices de facilité des items 55 à 72 répondus dans les sous-tests de 18 items et dans le test intégral.....	77
21	Analyse de la variance des indices de facilité des items 1 à 18 répondus dans les sous-tests de 36 items et dans le test intégral.....	78
22	Analyse de la variance des indices de facilité des items 55 à 72 répondus dans les sous-tests de 36 items et dans le test intégral.....	78
23	Les moyennes des indices de facilité des items 1 à 18 et 55 à 72 pour les échantillons de sujets des plans ECIS réel et pour l'échantillon de sujets du test à 72 items.....	79

INTRODUCTION

Cette recherche porte sur le *Multiple Matrix Sampling*, une technique de cueillette et d'analyse de données sur échantillons que nous traduisons ici par *Echantillonnage combiné des items et des sujets* (ECIS). L'introduction comprend les parties suivantes: la description de la technique ECIS, une présentation du contexte particulier dans lequel elle est utilisée dans notre recherche, un exposé du problème étudié et l'énoncé du but de la recherche.

LA TECHNIQUE

Depuis qu'on estime les paramètres de tests pour diverses populations, dans la majorité des cas, on a estimé les paramètres d'une population en constituant un échantillon représentatif de sujets auxquels on a fait subir un test. Les statistiques de cet échantillon servent ensuite d'estimation des paramètres (notamment la moyenne, la variance, les coefficients de consistance interne et de fiabilité, les moments d'ordre 3 et 4) et de leurs erreurs-types. Parfois, dans les cas de populations de petite taille, il arrive que l'on emploie tous leurs membres pour établir les paramètres d'un test.

L'échantillonnage combiné des items et des sujets (ECIS) implique, d'une part, la formation d'un échantillon représentatif des membres de la population et, d'autre part, la répartition par échantillonnage de tous les items du test dans un certain nombre de sous-tests. Une fois le nombre de sous-tests connu, l'échantillon de sujets est à son tour subdivisé en sous-échantillons, de sorte que chaque sujet ne répond qu'à un seul des sous-tests; cet appariement des sujets et des sous-tests se fait aussi aléatoirement. Une fois les données ainsi recueillies, les estimations des paramètres du test intégral sont calculées à partir des statistiques partielles de chacun des sous-tests suivant des formules développées par Hooke (1953) et Lord (1960). L'application de l'ECIS à l'estimation des paramètres d'un test pour une population visée comprend donc les étapes suivantes: 1°- la construction d'un certain nombre de sous-tests par échantillonnage aléatoire des items du test intégral, 2°- la sélection d'autant d'échantillons représentatifs de la population visée qu'il y a de sous-tests, 3°- la distribution aléatoire ces sous-tests aux échantillons, 4°- l'administration des sous-tests, 5°- la correction des sous-tests et, 6°- l'estimation des paramètres.

Sans parler de ses autres applications, la technique ECIS employée pour obtenir les estimations de paramètres offre plusieurs avantages pratiques; Shoemaker (1971c) mentionne notamment:

- la réduction des erreurs-types d'estimation de la moyenne, de la variance et du coefficient de consistance interne (KR 21)
- la réduction du temps de passation
- la réduction des coûts de correction
- l'ouverture de milieux habituellement réfractaires aux longues périodes de passation des tests
- un temps d'administration limité, l'augmentation du nombre d'items analysés
- l'estimation simultanée de paramètres de plusieurs tests
- une économie financière réalisée sur l'ensemble de l'opération.

Toutefois, la technique ECIS s'appuie sur un postulat fondamental concernant la performance des sujets à chaque item, lequel s'énonce ainsi: la réponse d'un individu à un item est indépendante du test où il se

trouve, en particulier des items qui l'accompagnent dans ce dernier. En d'autres mots, il faut que les deux seuls facteurs influençant la performance d'un individu à tel item soient, selon le modèle probabiliste développé par Rasch (1966), la connaissance ou l'habileté de l'individu et le degré de difficulté de l'item; plus le sujet est connaissant ou habile et plus l'item est facile, plus la probabilité d'une bonne réponse s'approche de 1. En estimant les paramètres par échantillonnage unique des sujets, on ne modifie ni la présentation des items dans le test ni les conditions contrôlables de passation. La technique ECIS change les deux à la fois; chaque sous-test devient pour ses items un contexte de présentation différent de celui du test intégral et, à certains égards, les conditions de passation des sous-tests diffèrent de celles du test original.

Bien que, selon Knapp (1973), au début des années 50, William Turnbull avançât intuitivement l'idée que la meilleure estimation de la moyenne d'un test s'obtiendrait en administrant un seul item à autant de sous-ensembles de la population visée qu'il y a d'items dans le tout et en calculant la moyenne des moyennes d'items, il fallut attendre Lord (1960) pour développer une théorie statistique convenable. Même le calcul de la moyenne dépasse vite la mesure de la calculatrice et demande, pour une exécution rapide, l'usage de l'ordinateur. Par la suite, la connaissance des formules d'estimation amena un développement rapide des études sur la technique ECIS. A ce jour, il existe de bons programmes informatiques permettant de tirer de l'ECIS les estimations des paramètres de tests ordinairement calculés par les évaluateurs. Voyons maintenant à quoi peut servir la technique et comment nous l'appliquerons dans le cadre de notre étude.

LE CONTEXTE D'UTILISATION DE LA TECHNIQUE ECIS

La technique ECIS peut servir à recueillir des données dans plusieurs contextes. Jusqu'ici, la technique ECIS a été utilisée pour recueillir des données: dans des recherches expérimentales (Piper et Shoemaker, 1973), dans l'évaluation de programmes (Montague, 1971), dans la normalisation de tests, dans la construction d'échelles d'attitudes (Bursack,

1969; Pugh, 1971) et d'inventaires des besoins (Misanchuk, 1978), dans les sondages d'opinions (Loadman, 1972; Burton et Remer, 1972), dans des évaluations nationales d'acquisition de connaissances (Cahen, Romberg et Zwirner, 1970 a et b).

Nous avons l'intention d'étudier la technique ECIS lorsqu'elle est utilisée comme technique de collecte et d'analyse de données pour la normalisation d'un test, normalisation étant pris ici dans le sens d'*investigation expérimentale* tel que dans la définition suivante:

> "*Le test est une épreuve normalisée par laquelle on compare le sujet à une ou plusieurs tables, dressées à partir des résultats d'une investigation expérimentale ayant porté sur un échantillonnage représentatif de sujets de même catégorie.*" (Vocabulaire de l'éducation au Québec, 1968, p. 52)

LE PROBLEME

A un test composé d'items, un répondant obtient un résultat d'ensemble que l'on identifie comme sa performance; celle-ci est faite de la somme de ses performances à chaque item. Cette performance d'un répondant à un item, est-elle influencée par autre chose que sa capacité de répondre? Cette capacité de répondre du sujet est-elle soumise à des influences externes qui peuvent sensiblement l'altérer?

On peut se demander notamment si la performance d'un répondant à un item donné est influencée par le nombre d'items dans le test, la position de l'item dans le test, sa performance aux autres items, la durée du test. L'ensemble de ces facteurs compose l'essentiel du contexte de la performance d'un répondant à un item donné.

Eu égard à la technique ECIS, Sirotnik (1970) définit l'influence du contexte en ces termes: "(...) *any tendency for the matrix sample estimates of matrix population parameters to be different from those obtained had the examinees in fact responded to the population of items.*" (Sirotnik, 1970, p. 199)

Quelle influence cette modification ECIS de contexte a-t-elle sur la performance des répondants? Les résultats des recherches sont fort peu concluants. Jacobs et Wildemann (1969) concluent que les conditions de passation particulières à la technique ECIS ne modifient pas sensiblement les estimés de la moyenne, de l'écart-type, des indices de facilité et du coefficient de Spearman-Brown. En ce qui concerne l'estimation de la moyenne, les recherches de Sirotnik (1970) et Novak (1974) corroborent celles de Jacobs et Wildemann (1969) mais demeurent énigmatiques quant à l'estimé de la variance et suggèrent d'autres travaux avant de conclure. Feldt et Forsyth (1974) démontrent que, pour leur test de mathématique, l'influence du contexte se manifeste significativement, tandis qu'il n'en est rien pour leur test d'anglais. Hill termine son rapport sur une note prudente: "(...) *matrix sampling can be pratically applied when care is taken to minimize violation of assumptions.*" (Hill, 1975, p. 9)

Le dernier mot n'est point encore dit quant à l'importance de l'influence du contexte ECIS sur l'estimation des paramètres d'un test.

LE BUT DE LA RECHERCHE

Les résultats des recherches colligées au chapitre suivant laissent croire que les modifications au contexte occasionnées par l'utilisation de la technique ECIS peuvent influencer les répondants pour que leur performance soit différente de celle qu'ils auraient en situation normale. Il n'est cependant pas encore démontré à notre satisfaction que les effets sont négligeables ou appréciables. C'est pourquoi nous voulons apporter à des questions précises des réponses empiriques fragmentaires qui contribueront à une meilleure compréhension de l'effet du contexte ECIS sur l'estimation des paramètres et la performance des répondants.

Cette recherche a pour but, lorsque la technique de collecte des données par échantillonnage combiné des items et des sujets est utilisée dans un contexte de normalisation d'un test d'intelligence, (1°) *d'étudier les estimations ainsi obtenues des quatre premiers moments et* (2°) *de comparer le rendement des répondants selon la position des items dans le test.*

En particulier, nous voulons que le modèle empirique nous fournisse des éléments de réponse aux questions suivantes:

- *Quelle influence les modifications ECIS au contexte ont-elle sur les estimations de la moyenne, de la variance et des moments d'ordre 3 et 4?*

- *Les modifications ECIS au contexte influencent-elles différemment le rendement des répondants aux items selon que ces derniers se trouvent au début ou à la fin du test intégral?*

LES LIMITES DE LA RECHERCHE

Plusieurs facteurs limiteront l'étendue des généralisations émanant des résultats de notre recherche:

LE TEST

Le choix du test arrête (1°) le genre (une épreuve d'intelligence), (2°) la durée de l'épreuve (le nombre d'items qu'il contient), (3°) le niveau de difficulté qu'il représente pour la population-cible, (4°) la distribution des items les uns par rapport aux autres (ordre croissant de difficulté ou désordre) et (5°) certaines conditions matérielles de l'épreuve (administration collective ou individuelle, sur papier à correction automatique et informatisée).

LA POPULATION

Le choix d'une population en arrête (1°) l'âge, (2°) la capacité moyenne de performance générale au test et (3°) le milieu d'origine tant scolaire que socio-économique.

LES ITEMS

Le choix des items en arrête (1°) le type (à choix multiple et à

réponse unique, bonne 1 ou fausse 0), (2º) le genre de problème (problème court ou long à répondre), (3º) le contenu (vocabulaire, arithmétique, observation), (4º) les indices de facilité, leur étendue et leur distribution et (5º) les indices d'efficacité, leur étendue et leur distribution.

LES PLANS ECIS

Le choix des plans ECIS arrête (1º) leur nombre (nous en avons trois), (2º) le rapport entre le nombre de sous-tests (t), le nombre d'items par sous-test (k) et, plus tard, le nombre de sujets répondant à chaque sous-test (n); quoique ce dernier puisse varier d'un sous-test à l'autre, ces trois valeurs représentent souvent un plan ECIS par le triplet (t, k, n), et (3º) les modifications de contexte conséquentes à tel plan.

Comme facteur restrictif de généralisation, on peut aussi inclure l'ensemble des décisions qui fixent à la fois le schéma expérimental et les conditions d'expérimentation. Il serait prématuré de les énoncer ici car ces décisions sont implicitement comprises dans la description de l'expérience.

En définitive, comme le travail repose sur l'emploi d'une technique de cueillette de données, une fois les caractéristiques et le milieu d'application déterminés, il faut accepter la limitation à la généralisation qui en découle. De fait, nos résultats empiriques s'interpréteront beaucoup à la lumière des résultats des études réalisées jusqu'ici tant dans le champ particulier à la technique ECIS que dans celui plus général de la docimologie.

Dans cette introduction, nous avons présenté la technique et son champ d'utilisation particulier à cette recherche, nous avons formulé les grandes questions auxquelles la recherche s'attaque et nous avons précisé les principales limites de la généralisation. Au chapitre suivant, nous exposerons les fondements théoriques qui sous-tendent notre démarche expérimentale.

CHAPITRE I

LES ECRITS

Pour présenter les résultats des études sur l'échantillonnage combiné des items et des sujets (ECIS), nous avons adopté la ligne de présentation suivante: tout d'abord, le bilan du groupe des recherches *ex post facto* où la technique ECIS a été simulée aux fins de normalisation; en second lieu, l'état des résultats des travaux *a priori* où la technique ECIS a vraiment servi à recueillir les données; ensuite, l'exposé des recherches où les auteurs ont plus particulièrement fouillé la question des effets que des modifications imposées au contexte pouvaient avoir sur la performance des répondants; et, finalement, un regard sur quelques travaux d'ordre général portant plus spécifiquement sur l'un ou l'autre des facteurs composant un contexte d'épreuve.

La recension des recherches *ex post facto*, soit celles qui simulent le modèle sans que les données aient véritablement été recueillies par la technique ECIS, conduit à la conclusion que les fondements statistiques de la technique ECIS sont fiables. Non seulement les études portent-elles sur la comparaison des estimations ECIS avec celles que l'échantillonnage unique des sujets donnait, mais elles vérifièrent aussi l'influence de diverses caractéristiques des plans ECIS, des tests et des populations.

Dans les recherches *a priori*, la technique d'échantillonnage combiné des items et des sujets est vraiment utilisée pour recueillir les données. Ces recherches ont surtout eu le mérite d'élargir le champ d'utilisation de la technique ECIS, notamment à l'évaluation des apprentissages scolaires et des attitudes. Au demeurant, les recherches *a priori* apportent peu d'éclairage sur les effets spécifiques du contexte ECIS sur l'estimation des paramètres.

Les recherches sur l'effet de contexte visent à estimer les paramètres d'un test pour une population donnée en appliquant simultanément la technique ECIS et l'administration du test intégral à un échantillon. Le bilan de ces études ne laisse douter de la pertinence de poursuivre des travaux touchant l'effet de contexte; les résultats connus sont loin de répondre à toutes les questions que les études soulèvent. Les autres recherches colligées approfondissent notre connaissance des composantes de ce contexte d'administration d'un test collectif, et par conséquent des modifications que l'application de la technique ECIS entraîne.

LES RECHERCHES EX POST FACTO

PRESENTATION

Pour être utile, la technique ECIS devrait s'appuyer sur un modèle de traitement statistique des données recueillies qui estime correctement les paramètres. C'est au cours de la période allant de 1965 à 1975 que fut réalisée la majeure partie des travaux de vérification du modèle statistique.

Lord commença en 1955 (Lord, 1955; Johnson et Lord, 1958 et Lord, 1962) par l'étude de l'estimé de la moyenne; puis le courant est entretenu surtout par Shoemaker (1970a; 1970b; 1970c; 1971a; 1971b; 1972a; 1973a), suivi de Barcikowski (1970a; 1970b; 1972; 1974; avec Terranova, 1966; avec Moy, 1973; avec Olsen, 1975; avec Upp, 1976). Nombre d'autres chercheurs se penchent sur divers problèmes posés par l'ECIS, mais d'une façon moins soutenue.

Désireux de vérifier avec la plus grande sûreté la validité des équations d'estimation, les chercheurs s'appuyèrent majoritairement sur un modèle expérimental qui leur assurait la connaissance des paramètres de la population-cible au test à l'étude. Ils éprouvèrent d'abord la technique ECIS en la simulant à partir de banques de données obtenues préalablement à l'occasion de la normalisation de tests par échantillonnage unique des répondants, d'où l'appellation de recherche *ex post facto*. Nous voulons ici montrer l'acquis de cette recherche et la fiabilité de

la théorie statistique qui sous-tend l'estimation de paramètres par la technique ECIS.

Le coup d'envol de la recherche *ex post facto* est donné par Lord et Johnson (1958) qui montrent que l'erreur-type de la moyenne d'un test de 11 items (en calcul différentiel) est plus petite lorsqu'elle est calculée en prenant le résultat de chaque étudiant à un seul item (appariement étudiant-item aléatoire) plutôt qu'en calculant la moyenne au test intégral d'un échantillon d'étudiants. Fort des promesses de cette étude, Lord publia en 1962, les résultats de la première étude véritable de l'ECIS, résultats concernant la moyenne et la variance ainsi que la distribution de probabilités. Pour réaliser son étude, il utilisa *ex post facto* les scores de 1 000 sujets à un test de 70 items. Les estimations ECIS, aussi bien de la moyenne que de la variance, se révélèrent plus précises et fidèles que leurs correspondantes par échantillonnage unique des sujets (EUS). Pour estimer la distribution de probabilités, il appliqua la distribution hypergéométrique négative aux échantillons EUS et ECIS. Encore là, les distributions ECIS furent toutes plus précises que leurs correspondantes EUS. La conclusion fut que la technique ECIS pouvait servir avec justesse à l'estimation des paramètres.

Comme il avait procédé par échantillonnage non exhaustif des items, seulement 45 des 70 apparaissaient dans les sous-tests, dont 18 dans plusieurs sous-tests. Lord, reconnaissant là une faiblesse de son étude, suggéra qu'une autre recherche reproduise son expérience mais en composant les sous-tests par échantillonnage exhaustif. Plumlee (1964) suivit la suggestion de Lord et appliqua *ex post facto*, elle aussi, la technique ECIS à un test de 30 items (en ponctuation) donné à 200 candidats, à des postes de catégorie "bureau". Elle appliqua aussi la technique EUS pour pouvoir comparer la précision des estimations ECIS à leurs correspondantes EUS.

Pour la moyenne, elle confirma les résultats obtenus par Lord, mais pour la variance, les estimations EUS se révélèrent plus précises que leurs correspondantes ECIS. Malgré les faiblesses d'estimation de l'écart-type, Plumlee conclut que ses résultats étaient satisfaisants et n'infirmaient pas l'hypothèse de justesse de l'estimation par la technique ECIS.

Barcikowski et Terranova (1966) voulaient établir le plus petit nombre d'items qu'il faut servir à chaque sujet pour obtenir des estimés fidèles de la moyenne et de la variance. Ils appuyèrent leur étude sur des données simulées par une méthode informatisée dite "Monte Carlo". Bien qu'ils ne résolurent pas complètement leur problème, ils purent confirmer, pour chacune des trois combinaisons d'étendues des indices de facilité et des indices d'efficacité, qu'il existait une taille des sous-tests fournissant des statistiques ECIS plus précises que leurs correspondantes EUS.

Ayant remarqué que les sous-tests comprenaient 10% d'items du test intégral dans les recherches de Lord (1967) et Plumlee (1964), Cook et Stufflebeam (1967) étudièrent la variable nombre d'items par sous-test en composant des sous-tests à 10%, 25%, 33% et 50% du total d'items. Ils procédèrent par comparaison avec la technique EUS à partir d'un test de 115 items ayant déjà été soumis à 1 239 étudiants.

Les résultats confirmèrent ceux de Lord (1962) et de Plumlee (1964) concernant la moyenne et ceux de Lord (1962) concernant la variance. En plus, ayant estimé les coefficients 20 et 21 de Kuder-Richardson ainsi que le Spearman-Brown, les auteurs obtinrent des statistiques ECIS plus exactes que leurs correspondantes EUS. Quant à la taille des sous-tests et des échantillons de sujets, ils conclurent que ces variables influençaient peu la précision des estimations.

Knapp, dans une étude rapportée en 1968, se préoccupa surtout de la taille de l'échantillon total ECIS, de la taille des sous-tests et de l'appariement des items aux sujets. Il se servit de la méthode du plan à sections incomplètes équilibrées (*balanced incomplete block design*) pour déterminer le plan ECIS; une fois le nombre d'items par sous-test déterminé, cette méthode calcule à la fois la taille de l'échantillon total de sujets et celle des sous-échantillons. Il fit son étude parallèlement sur trois populations d'items, la première de 29 soumise à 4 060 étudiants du secondaire, les deux autres distinctes, chacune de 43 soumise à 903 autres étudiants, également de niveau secondaire. Il estima la moyenne, la variance et le coefficient 20 de Kuder-Richardson.

Knapp jugea les statistiques ECIS tout à fait satisfaisantes. Il remarqua que le fait d'ajouter des items aux sous-tests n'améliorait pas tellement la précision. Toutefois, dans l'ensemble, les estimations EUS étaient légèrement plus précises; selon lui, ce gain de précision ne justifie pas le surcroît de travail et de temps occasionné.

Comme suite au travail de Cook et Stufflebeam (1967), Shoemaker (1970a) s'interrogea à son tour sur le nombre idéal de sous-tests, leur taille, la taille des sous-échantillons de sujets. Il créa 20 plans ECIS à partir d'un test de 150 items déjà administré à 810 étudiants. Pour chaque plan, il estima la moyenne, l'écart-type, l'erreur-type dans chaque cas et le coefficient 21 de Kuder-Richardson. En plus, il estima la distribution de probabilité selon le modèle hypergéométrique négatif.

Des résultats obtenus par Shoemaker, il appert que c'est le nombre total d'observations, une observation étant une réponse d'un sujet à un item, qui influence davantage la précision de l'estimation. Notamment, il rapporta que pour obtenir des résultats valides, il lui avait fallu prendre 1,23% du total des observations, soit 1 500 sur 121 500. Lord (1962) et Plumlee (1964) en prirent chacun 10% et Cook et Stufflebeam (1967) entre 9,18% et 49,45%.

En construisant les plans ECIS, Shoemaker en obtint sept qui contenaient 50% ou moins des items du test intégral; ce fait entraîna des écarts plus importants de l'estimation de la distribution de probabilités.

Désireux de connaître l'effet de la distribution de probabilités sur la précision de l'estimation par l'ECIS, Shoemaker (1970b) compara les statistiques pour une distribution négativement biaisée à celles d'une distribution normale. Concurremment, il étudia les effets de l'omission d'items. Il fonda son étude sur les réponses de 1 031 sujets à un test de 20 items.

Ses résultats confirmèrent l'hypothèse que la précision de l'estimation de la moyenne et de l'écart-type est surtout fonction du nombre d'observations. L'estimation ne semble pas sensiblement influencée par l'asy-

métrie de la distribution de probabilités. Plus le nombre d'observations est élevé, moins l'omission d'items a d'effets sur la précision; il ne s'est manifesté aucune variation corrélative entre les erreurs-types et le nombre d'items omis.

L'importance du nombre d'observations est aussi démontrée par Hambleton, Rovinelli et Gorth (1971) qui, en plus, confirmèrent le peu d'influence qu'a la courbe de la distribution de probabilités sur l'estimation ECIS. Maintenant le nombre d'observations constant, ils déterminèrent qu'il valait mieux, pour accroître la précision, réduire la taille des sous-tests et augmenter celle des sous-échantillons de sujets que de faire le contraire. Pour leur étude, ils utilisèrent le modèle de simulation informatique "Monte Carlo".

En 1971, Shoemaker (1971b) entreprit une étude corollaire de ses deux précédentes (1970a et b) dans laquelle il tentait de répondre à la question: pour un test déterminé, tous les plans ECIS ayant le même nombre d'observations ont-ils, à toutes fins utiles, des erreurs-types d'estimation de la moyenne et de la variance comparables?

Partant d'une banque de données, il construisit deux distributions normatives, l'une normale et l'autre biaisée négativement; pour ce faire, il fit appel à la technique de simulation informatique "Monte Carlo". Shoemaker créa trois plans ECIS et trois nombres d'observations pour chaque distribution.

Cette recherche confirma l'importance du nombre d'observations; la précision s'accroît avec le nombre d'observations. Tout comme Hambleton, Rovinelli et Gorth (1971), il constata l'influence significative du nombre de sous-tests; la précision croît avec le nombre de sous-tests. Pour la première fois, cette recherche révéla que l'étendue des indices d'efficacité aurait des effets sur l'estimation. Finalement, encore une fois, la courbe de distribution, normale ou asymétrique, n'avait relativement pas d'incidence sur l'estimation.

Utilisant une analyse de variance bidimensionnelle pour éprouver ses

résultats, Newell (1971) confirma les résultats précédents selon lesquels la moyenne et la variance peuvent être estimées avec une précision satisfaisante par l'ECIS. Il montra aussi que la technique ECIS se prête à l'analyse de la variance, et que le type d'échantillonnage des items (stratifié ou non) n'influence pas l'estimation.

Barcikowski (1972) reprit l'étude de l'estimation ECIS de la moyenne et de la variance mais en contrôlant les variables suivantes: le nombre d'items dans le test intégral, l'étendue des indices de facilité des items et l'étendue de leurs indices d'efficacité. Comme en 1966, il simula des données par le procédé informatique "Monte Carlo". Il tira ses conclusions de l'analyse des écarts entre les statistiques ECIS et EUS tant pour la moyenne que pour la variance.

Quelles que soient les conditions, la technique ECIS apparut plus précise que la technique EUS pour estimer la moyenne. Quant à la variance, pour les tests ayant des r_{bis} homogènes, l'estimation ECIS fut plus précise que la correspondante EUS, tandis que pour les tests à r_{bis} hétérogènes, aucune tendance ne se manifesta. Par ailleurs, parmi les tests à 97, 150 ou 194 items, ce furent les tests à 97 items qui obtinrent l'estimation la plus précise de la moyenne et de la variance. Donc, Barcikowski (1972) démontra que l'étendue des r_{bis} et le nombre d'items sont des facteurs non négligeables influençant l'estimation ECIS de paramètres.

Dans une autre étude *ex post facto*, Shoemaker (1972a) fit varier le coefficient de fidélité 20 de Kuder-Richardson (KR 20), la variance (σ^2) des indices de facilité et le degré d'asymétrie (courbe normale, courbe très négativement biaisée). Il manipula sept plans ECIS distincts sur un test de 40 items.

Ces résultats confirmèrent essentiellement que le nombre d'observations est inversement proportionnel à l'erreur-type d'estimation d'un paramètre. Parallèlement, il sembla que l'augmentation du nombre de sujets passant chaque sous-test était la façon la moins efficace de réduire l'erreur-type d'estimation; dans les cas de distribution normale, ce serait la taille des sous-tests qui donnerait le meilleur résultat, alors que

pour les distributions négativement biaisées, on devrait choisir un plan ECIS à nombre élevé.

D'autre part, pour un KR 20, un nombre d'items et un plan ECIS déterminés, la croissance de σ_p^2 entraînerait une augmentation de l'erreur-type d'estimation de la moyenne et une diminution de celle de la variance. Le même phénomène se produisit dans les cas où σ_p^2, le nombre d'items et le plan ECIS étaient déterminés, mais où KR 20 variait.

Autre révélation de l'étude de Shoemaker, elle fournit une preuve supplémentaire que, lorsque le produit du nombre de sous-tests (t) par leur taille (n) est supérieur au nombre total d'items (k) d'une part, t k devrait être un multiple entier de k et, d'autre part, tous les items devraient apparaître dans le même nombre (tk/k) de sous-tests.

Kleinke (1972b; 1973), remarquant que seuls Lord (1962) et Cook et Stufflebeam (1967) avaient abordé l'estimation de la distribution de probabilités, résolut d'étudier d'abord l'estimation ECIS du score au test intégral et de comparer ensuite des modèles d'approximation de la distribution de probabilités. Quelle que soit la distribution normative, le modèle hypergéométrique négatif fournit une courbe approximative de la distribution normative, jugée pleinement satisfaisante. L'équation de prédiction linéaire dont il se servit pour estimer les scores au test intégral, quoique présentant quelques faiblesses, donna une estimation en général acceptable et sans biais systématique.

En 1973, Shoemaker (1973a) entreprit l'étude de deux problèmes: le premier étant la façon de constituer les sous-tests, le second, l'emploi de la méthode statistique dite *jackniſe* pour calculer les erreurs-types d'estimation. Pour étudier les façons de former les sous-tests, il compara l'échantillonnage non exhaustif des items à l'échantillonnage exhaustif. Au moment de construire sa distribution *ex post facto*, Shoemaker tint compte des variables suivantes: le nombre d'items par test, la variance ces indices de facilité, le coefficient de fidélité, le degré d'asymétrie et le type de plan ECIS.

En substance, les résultats s'expriment ainsi: (1°) la variation du coefficient de fidélité a peu d'incidence sur les erreurs-types d'estimation; (2°) l'échantillonnage non exhaustif réduit l'erreur-type d'estimation de la moyenne mais cette réduction devient progressivement moins significative pour les deuxième, troisième et quatrième moments; (3°) pour une combinaison t k déterminée, l'augmentation de t (nombre de sous-tests) produit une diminution de l'erreur-type d'estimation de la moyenne, tandis que c'est l'accroissement de k qui réduit l'erreur-type d'estimation des autres moments; (4°) quoique avec une légère tendance à surévaluer l'erreur-type d'estimation de la moyenne, la méthode *jacknife* estime bien les erreurs-types.

Brandenburg et Forsyth consacrèrent (1973a et b) deux études aux modèles d'approximation des distributions de probabilités dans un contexte ECIS. Bien qu'ils apportèrent quelques réserves au modèle hypergéométrique négatif, ils conclurent que son ajustement à la distribution normative était tout à fait acceptable.

Non satisfait des résultats antérieurs, Barcikowski, aidé de Moy (1973), tenta à nouveau de déterminer, dans des conditions données, quelle combinaison k (taille des sous-tests) par n (taille des sous-échantillons de sujets) donne les estimés les plus précis. Plus spécifiquement, ils axèrent leur étude sur les deux questions suivantes: 1) étant donné certaines étendues de r_{bis} connues, quelle combinaison items-sujets par sous-test estimera plus précisément la moyenne et la variance? 2) laquelle des deux façons de constituer des sous-tests, l'échantillonnage exhaustif simple ou l'échantillonnage exhaustif avec remise entre deux sous-tests, donnera un plus haut degré de précision des estimés? Les auteurs firent encore appel à la méthode informatique "Monte Carlo" pour composer les distributions expérimentales.

Quelles que soient les valeurs des r_{bis} et leur étendue, l'analyse des données révéla que l'estimation optimum de la moyenne provient de plans ECIS à plusieurs sous-tests comptant peu d'items, résultat conforme à ceux de Barcikowski (1970b) et de Shoemaker (1971b). Par ailleurs, l'estimation de la variance est fortement influencée par la nature des r_{bis};

ainsi, lorsque les r_{bis} sont élevés, c'est aussi le grand nombre de sous-tests à quelques items qui donne de la précision aux plans ECIS, mais lorsqu'ils sont faibles, un petit nombre de sous-tests à plusieurs items fournit une estimation plus précise. Les tests composés d'items à r_{bis} moyens semblaient s'accommoder mieux de plusieurs sous-tests à peu d'items.

Quant aux types d'échantillonnage, les auteurs recommandèrent d'emblée l'échantillonnage exhaustif simple. Cette recommandation s'accompagnait d'une réserve, à savoir que le nombre de sous-tests à peu d'items (cas-type: 500 sous-tests de 6 items) et la précision de l'estimation de la moyenne augmentaient parallèlement; au-delà d'une limite, le gain devient minime comparé aux difficultés d'exécution de tels plans.

Dernière connue en date, l'étude *ex post facto* de Scheetz (1976) couvrit plusieurs des questions auxquelles les auteurs précédents apportèrent des réponses partielles. Parmi les conclusions de son travail, signalons qu'étant donné un nombre d'observations fixes, si tk = K, alors ce furent les plans ECIS à t et k moyens qui donnèrent les statistiques les plus précises. Cependant si tk = 2K, alors ce furent les plans ECIS où t est beaucoup plus grand que k qui réduisirent l'erreur-type d'estimation de la moyenne. Au total, sur les variables indépendantes t et k, cette recherche apporta peu d'éclairage nouveau.

Dans la deuxième partie de son travail, Scheetz étudia les effets de l'échantillonnage stratifié des items pour la composition des sous-tests sur la précision de l'estimation. Ayant utilisé deux tests, l'un de vocabulaire et l'autre en mathématique, il constata, pour le test de vocabulaire, une estimation plus précise de la variance lorsque les items étaient distribués par échantillonnage stratifié selon les indices de facilité. Pour le reste, Scheetz confirma Myerberg (1975) qui n'avait pu enregistrer de différence perceptible sur les erreurs-types de la moyenne et de la variance entre un plan à échantillonnage simple et un plan à échantillonnage stratifié des items selon leurs indices de facilité. En stratifiant selon les indices d'efficacité, il obtint une diminution de l'erreur-type de la moyenne.

Reprenant, dans la troisième partie de son étude, la comparaison des plans ECIS développés avec le modèle à sections équilibrées incomplètes (*partially balanced incomplete block designs*) aux plans ECIS construits par échantillonnage exhaustif simple, Scheetz découvrit que les premiers seraient supérieurs au second pour estimer la moyenne, mais que ce serait l'inverse dans le cas de la variance.

BILAN

Le bilan de la recherche *ex post facto* se trace en deux volets caractérisés par le choix des variables indépendantes. Le premier volet regroupe ces recherches conduites entre 1958 et 1968 où les auteurs ont comparé les statistiques ECIS à leurs correspondantes EUS reconnues jusque-là fiables. La précision de ces statistiques constituait la variable dépendante alors que la technique de cueillette de données servait de variable indépendante. On comparait les techniques de l'échantillonnage combiné des items et des sujets (ECIS) et de l'échantillonnage unique des sujets (EUS), toujours *ex post facto*, sur des populations définies afin de connaître les paramètres qui servaient de critères de comparaison.

Depuis 1970, et ce jusqu'à nos jours, il semble que les chercheurs aient accepté le fait que les statistiques ECIS sont au moins aussi précises et fiables que leurs correspondantes EUS et ont commencé à étudier les facteurs influençant la précision d'estimation de la technique ECIS. Dans cette série d'études, on fit varier les diverses caractéristiques des plans ECIS, des tests et des populations à l'étude: la précision demeurant la variable dépendante.

Le premier volet de recherche démontre, de toute évidence, que la théorie mathématico-statistique de l'échantillonnage combiné des items et des sujets donne des statistiques précises et fiables des paramètres (moyenne-variance) d'une population à un test. Même si la preuve est plus discutable, la technique ECIS peut aussi estimer les indices KR 20 et KR 21, Spearman-Brown et la distribution de probabilités. Le sommaire des recherches composant le premier volet est présenté au tableau 1; on y retrouve les principaux résultats tirés de comparaisons entre les statistiques ECIS et les statistiques EUS.

TABLEAU 1

Sommaire des résultats des recherches *ex post facto* comparant la technique ECIS à la technique EUS pour estimer les paramètres

RECHERCHE		ESTIMATION DE		
Auteurs	Année	μ	σ^2 ou σ	Autres
Lord et Johnson	1958	$S_{\bar{x}}$ ECIS < $S_{\bar{x}}$ EUS		
Lord	1962	$S_{\bar{x}}$ ECIS < $S_{\bar{x}}$ EUS	S_{S^2} ECIS < S_{X^2} EUS	Distribution de probabilités: ECIS plus précis.
Plumlee	1964	ECIS plus précis	EUS plus précis	
Barcikowski et Terranova	1966	ECIS plus précis	ECIS plus précis	
Cook et Stufflebeam	1967	ECIS plus précis	ECIS plus précis	Indices KR 20, KR 21 et Spearman-Brown: ECIS plus précis.
Knapp	1968	EUS plus précis	EUS plus précis	

$S_{\bar{x}}$ EUS Erreur-type de l'estimateur de la moyenne par échantillonnage unique des sujets

S_{S^2} EUS Erreur-type de l'estimateur de la variance par échantillonnage unique des sujets

μ Moment

σ^2 Variance

σ Ecart-type

$S_{\bar{x}}$ ECIS Erreur-type de l'estimateur de la moyenne par échantillonnage combiné des items et des sujets

S_{S^2} ECIS Erreur-type de l'estimateur de la variance par échantillonnage combiné des items et des sujets

Bien que les recherches *ex post facto* du deuxième volet aient fait intervenir de nouvelles variables indépendantes, elles n'en ont pas moins, par ce biais, continué à confirmer les principaux résultats des recherches groupés dans le premier volet. C'est avec les travaux de Shoemaker, publiés en 1970, que débute cette vague qui s'étend jusqu'à 1976; depuis ce temps, aucun travail pouvant s'inscrire dans cette lignée ne semble avoir été publié sauf, peut-être, Sachar (1980) [1] pour plus tard.

Ces recherches ont contribué à écarter quelques inquiétudes quant à l'influence de certains facteurs. Ainsi, au moment de décider du rapport entre la taille des sous-tests, le nombre de sous-tests et la taille des sous-échantillons de sujets, il n'y aurait lieu de se préoccuper ni de l'allure de la distribution de probabilités (Shoemaker, 1970b, 1971b; Hambleton *et al.*, 1971), ni du type d'échantillonnage des items adopté pour la composition des sous-tests (Newell, 1971; Shoemaker, 1973a; Moy et Barcikowski, 1973; Myerberg, 1975; Scheetz, 1976). Dans les cas où plusieurs paramètres sont estimés, les résultats, avec certaines réserves, suggèrent à l'utilisateur de l'ECIS de prendre en compte le nombre total d'observations plutôt que le rapport t : k, le nombre de sous-tests plutôt que leur taille (Cook et Stufflebeam, 1967; Shoemaker, 1970a, 1971b, 1972, 1973a; Hambleton *et al.*, 1971; Scheetz, 1976). Enfin, Newell (1971) montra que l'ECIS se prête à l'analyse de variance.

Finalement, les recherches *ex post facto* mirent en lumière l'influence de certains facteurs indépendants et de leurs interactions sur la précision des estimés ECIS en regard de la précision de l'estimation. Barcikowski et Terranova (1966) mirent en évidence une relation entre, d'une part, une combinaison de l'étendue des indices de facilité et de celle des indices d'efficacité des items et, d'autre part, la taille des sous-tests. L'étendue des r_{bis} ainsi que leurs valeurs se révélèrent source

---

1 En septembre 1980, la revue Educational and Psychological Measurement publiera un article de J. Sachar intitulé "Estimating total-test scores from partial scores in a matrix sampling design".

d'influence sur la précision de l'estimation (Barcikowski, 1970b, 1972; Shoemaker, 1971b, 1972; Moy et Barcikowski, 1973); ces recherches montrèrent, par ailleurs, que la variation de t ou k produisait des effets contraires sur la précision de l'estimation de la moyenne et de la variance.

En résumé, les recherches *ex post facto* rassurent l'utilisateur sur la robustesse et la justesse de l'estimation ECIS, indiquent la méthode à suivre pour bien planifier une opération d'échantillonnage combiné des items et des sujets et met en garde contre les écueils les plus sérieux. Tout cela étant valide pourvu que se vérifie le postulat d'absence d'effet du contexte sur la performance des répondants.

LES RECHERCHES A PRIORI

PRESENTATION

Les recherches *ex post facto* avaient dans leur démarche le mérite de fournir les paramètres de la population-cible. Cependant, elles présentaient l'inconvénient majeur de ne pas soumettre la technique ECIS à l'épreuve de l'usage réel. Avec Owens et Stufflebeam (1969) débuta un courant d'études où les données de normalisation étaient recueillies par l'application de la technique ECIS; en procédant ainsi, on pouvait ou bien altérer le mode d'emploi et obtenir encore les paramètres de la population ou bien respecter la technique ECIS et perdre les paramètres comme critères de mesure des statistiques ECIS.

Owens et Stufflebeam (1969) furent les premiers à cueillir leurs données selon un plan d'échantillonnage combiné des items et des sujets déterminé d'avance. Ils expérimentèrent simultanément trois plans ECIS (17 sous-tests/3 items par sous-test/22 sujets par sous-test; 8/6/47; 4/12/14) à partir d'un test de vocabulaire de 50 items d'une durée de quinze minutes, administré à des élèves de 4e année répartis en deux groupes, l'un de faible et l'autre de forte aptitude.

Ils administrèrent tous les items à tous les sujets, mais ceux qui fu-

rent désignés à un des sous-tests ECIS répondirent d'abord aux items du sous-test sur une première page avant d'entreprendre les autres items sur les pages suivantes.

Quoique plus précise pour le groupe des élèves faibles que pour celui des forts, l'estimation ECIS de la moyenne était satisfaisante et comparable à l'estimation EUS. Pour la population de faible aptitude, l'estimation de la moyenne et de l'écart-type se précisa à mesure que la taille des sous-tests s'accroissait; c'était la réduction de la taille qui avait le même effet pour la population forte. Owens et Stufflebeam se gardèrent bien d'en tirer quelque conclusion, vu le peu de plans ECIS compris dans leur expérience.

Parmi les réserves que l'on peut exprimer sur cette étude, il faut certainement attirer l'attention sur le fait que la durée de passation était relativement courte (15 minutes) et que, par surcroît, elle n'était aucunement modifiée puisque les auteurs administrèrent les 50 items à tous les sujets. En composant les sous-tests, les auteurs durent parfois omettre quelques items, parfois en inclure dans plus d'un sous-test, dérogeant ainsi à la consistance de leurs plans ECIS. Cette recherche, compte tenu de ces limitations, permet néanmoins de constater que l'ordre de présentation des items n'a pas d'effet sur la précision de l'estimation de la moyenne et de la variance.

La même année, Bursack (1969) appliqua l'ECIS à la validation d'une échelle d'attitudes face à la mathématique chez les étudiants entrant à l'université. Il compara les résultats de l'administration du questionnaire intégral à ceux de l'administration de huit sous-questionnaires. De son expérimentation, il conclut que la technique ECIS avait provoqué des réponses plus valides de la part des sujets que celles obtenues en appliquant tout l'instrument. Cette recherche avait surtout le mérite d'ouvrir un nouveau champ d'application pour la technique ECIS.

La première de deux études assez extensives conduites par Cahen, Romberg et Zwirner (1970a) comparait les moyennes de 81 écoles à deux versions parallèles d'un test de rendement en arithmétique de 50 items;

la version B du test fut administrée le premier jour selon un plan ECIS, alors que la version A le fut intégralement le lendemain. Soixante-et-onze moyennes ECIS d'école étaient supérieures à leurs correspondantes au test parallèle administré intégralement, état de choses nullement attribuable au hasard. Par contre, l'ordre des écoles établi par ECIS correspond à celui tiré du test version A (corrélation de ,900).

La méthodologie suivie par Cahen et ses deux collègues demande quelques mises au point: 1o) leurs sous-tests comprenaient d'autres items que ceux qui servaient à l'étude; 2o) le test était chronométré, contrairement à une réserve exprimée par Lord (1962); 3o) le nombre de sujets variait d'une école à l'autre; 4o) dans 22 écoles, certains sous-tests ne furent pas administrés. Malgré tout, c'est la première recherche où l'on remarque une tendance systématique à la surestimation ECIS de la moyenne.

En complément à la précédente, Cahen et ses deux collègues (1970b) répliquèrent avec une seconde étude s'appuyant sur une stratégie expérimentale différente. Cette fois, ils travaillent avec des étudiants de 12e année et un test de 24 items en mathématique considéré difficile. Dans chaque école ils répartissent les sujets en deux groupes; le premier jour, l'un reçut le test intégral, tandis que l'autre passa un test placebo. Le lendemain, les deux groupes de sujets passèrent l'un des sous-tests ECIS, le même plan ECIS répété. Chacune des écoles comptait au moins cinq sujets dans deux groupes.

Cette recherche établit sans contredit la validité de l'estimation ECIS de la moyenne et de la variance des écoles. Elle confirma un résultat corollaire de la précédente (Cahen *et al.*, 1970a) indiquant une relation entre l'écart de la moyenne ECIS à la moyenne du test intégral et la racine carrée du nombre de sujets dans l'école, constatation se rattachant à celle qui lie la précision de l'estimation au nombre d'observations (Cook et Stufflebeam, 1967; Shoemaker, 1970a, 1971b, 1972c, 1973a; Hambleton *et al.*, 1971; Scheetz, 1976).

Encore ici, l'étude de Cahen et ses collègues appelle quelques réserves: 1o) ils utilisent encore un test chronométré; 2o) dans trois écoles, tous les sous-tests ECIS n'ont pas été administrés à au moins un

sujet; 3°) leurs analyses portent surtout sur la moyenne.

Montague (1971) s'intéressait à l'emploi de la technique ECIS pour évaluer un programme d'enseignement. En particulier, elle se demandait comment extraire de la banque les items qui composent les tests lorsque ces derniers sont à leur tour partagés en sous-tests ECIS. Elle composa concurremment plusieurs méthodes d'échantillonnage des items et plusieurs plans ECIS. Elle utilisa des items sur les habiletés arithmétiques fondamentales aux niveaux des 2e et 3e années. Dans un cadre expérimental assez complexe, pour l'évaluation d'un programme d'enseignement, l'emploi de la technique ECIS a fourni des résultats comparables dans toutes les situations expérimentées. D'où la conclusion générale de l'auteur: *"The feasability of concurrently and randomly sampling examinees and items in order to obtain group data generalizable to a universe of examinees and to universe of items has been strongly demonstrated in the present study"* (Montague, 1971, p. 181). Bien que Montague n'ait pas constaté d'influence du contexte ECIS sur le rendement des répondants, ayant comparé des plans ECIS en regard des façons de les constituer, son plan expérimental annulait l'effet de contexte parce que, s'il existait, il apparaissait dans tous les traitements.

Poggio et Glasnapp (1973) appliquèrent l'échantillonnage des items à la mesure du rendement scolaire. Ils conçurent trois examens périodiques, chacun en six versions parallèles, différentes sur la moitié des items qui les composaient. Chaque étudiant passait une version de chacun des tests dans laquelle il répondait à dix items communs aux six versions et à dix autres propres à sa version du test périodique. Les versions se distinguaient par les dix derniers items choisis au hasard dans une banque de dix. Les auteurs voulaient connaître les capacités de prédiction des blocs d'items communs d'une part, et d'items propres d'autre part.

Une première analyse révéla à Poggio et Glasnapp, pour un même examen périodique, une faible corrélation entre la moyenne aux items communs et celle aux items échantillonnés, quoique leur différence ne fût pas significative. Les différentes analyses effectuées, si elles assurèrent une certaine équivalence entre les résultats aux items communs et ceux

aux items échantillonnés, révélèrent néanmoins une meilleure possibilité de prédiction de ces derniers.

Rothrock (1976) et Michael, Forester (1976) et Michael, dans la foulée de Poggio et Glasnapp (1973), empruntèrent le schéma expérimental de ces derniers, mais avec moins de succès. Leurs conclusions s'inscrivent à l'enseigne de la prudence: c'est surtout la présence de trois différences de moyennes sur six, statistiquement significatives qui les inquiéta. Hormis ces écarts, les autres analyses amenèrent les auteurs à conclure que les résultats tirés des dix items communs équivalent à ceux des items échantillonnés.

BILAN

A l'exception de celle d'Owens et Stufflebeam (1969) où ces derniers modifièrent d'ailleurs la technique ECIS, les recherches *a priori* inventorièrent plutôt de nouvelles applications de l'échantillonnage combiné des items et des sujets que sa valeur propre dans l'estimation de paramètres. Cahen, Romberg et Zwirner (1970a et b) mirent à profit la technique ECIS dans le contexte vaste de l'appréciation, à l'échelle nationale (Etat américain), du niveau d'acquisition de connaissances de certaines matières pour toute une population écolière. Avant eux, Bursack (1969) avait échantillonné avec succès les items d'un questionnaire d'attitudes et, l'année suivante, Montague (1971) testa l'applicabilité de la technique ECIS à l'évaluation de programmes. Poggio et Glasnapp (1973a et b) suivis de Rothrock (1976), Forester (1976) et Michael appliquèrent avec un succès mitigé l'échantillonnage d'items dans la composition d'examens périodiques.

En définitive, cette voie de recherche s'écarte légèrement des problèmes d'estimation ECIS de paramètres pour nous faire voir la diversité d'application de la technique. L'influence du contexte n'étant pas leur objet, les recherches *a priori* apportent peu d'éclairage sur les effets spécifiques du contexte ECIS sur l'estimation des paramètres.

LES RECHERCHES SUR L'EFFET DE CONTEXTE

PRESENTATION

"It is of course true that a student cannot be expected to behave in exactly the same way when taking a single item as he would if this item were part of a long test" (Johnson et Lord, 1958, p. 329). Soupçonnant un certain effet du contexte ECIS, Lord écrivait ultérieurement:

> *"Any practical application of the item - sampling method thus involves the further assumption that the examinees' performance on the item is not too greatly affected by the context in which they are administered. Statistical theory can hardly be expected to dispense with the need for this assumption. Presumably the assumption will be tolerable in some practical situations but not in others; it must therefore be justified empirically for the actual practical situation in which the theory is to be applied."* (Lord, 1962, p. 263)

Très tôt, les chercheurs se sont intéressés à savoir si les modifications que subissaient et le test dans sa présentation et les conditions de passation, telles la durée et l'importance de la tâche, auraient des conséquences appréciables sur la précision et la justesse de l'estimation. Il fallut d'abord s'assurer que la technique d'échantillonnage combiné des items et des sujets (ECIS) puisse concurrencer, quant à la justesse et à la précision de l'estimation, la technique d'échantillonnage unique des sujets (EUS); pour ce faire, la recherche *ex post facto* convenait admirablement bien. Mais toutes les études présentées jusqu'ici, ou bien écartent le problème de l'effet de contexte ou bien postulent l'absence d'effet notable sur le rendement des répondants. Néanmoins, quelques chercheurs se sont penchés sur l'effet de contexte; nous ferons maintenant état de ces travaux.

Dans la première recherche où le schéma expérimental permettait de déceler l'influence du contexte, les auteurs Jacobs et Wildemann (1969) comparèrent les estimations de la moyenne, des indices de facilité et du coefficient de Spearman-Brown obtenus par l'utilisation réelle de la technique ECIS à celles obtenues par sa simulation. Dans le cadre d'un

cours universitaire, les chercheurs préparèrent un examen de 100 items soumis à 105 étudiants, tandis que 95 autres étaient assignés à l'un des 19 sous-tests des quatre plans ECIS qu'ils avaient créés, soit les plans ECIS (10 sous-tests, 10 items/sous-test, 5 sujets/sous-test), (4,25,5), (3,33,5) et (2,50,5), tous les sous-tests de chaque plan étant, au niveau des items, mutuellement exclusifs. Des données de l'examen intégral, autant de plans ECIS comparables furent constitués par simulation.

Sur 19 paires de moyennes, deux seulement montrèrent un écart significatif à 5%; par contre, toutes les paires de variances sont comparables. L'estimation ECIS réel ne s'écartait pas sensiblement de l'estimation déduite des résultats des 105 sujets au test intégral. Choisissant aléatoirement 15 items, Jacobs et Wildemann calculèrent leurs indices de facilité avec les données des plans ECIS et obtinrent des estimations jugées satisfaisantes. On nota que l'application de la technique ECIS montre une tendance quasi systématique à la surestimation de la moyenne et de l'écart-type.

Sirotnik (1970) voulait aussi comparer les estimations ECIS réel à celles obtenues par simulation et, en plus, aux paramètres mêmes de la population. Il travailla en milieu universitaire avec trois tests: l'un de vocabulaire V (30 items, 30 minutes), un second d'arithmétique M (30 items, 60 minutes) et un troisième, un questionnaire d'attitudes A (30 items, 15 minutes). Le premier jour, la moitié des sujets (90) passèrent le test V intégral et l'autre moitié, successivement, un des sous-tests V, M, A (ECIS) et le test V intégral. Le deuxième jour, tous les sujets passèrent le test M entier et, le jour suivant, le questionnaire A entier. Après coup, il simula ECIS par échantillonnage aléatoire des réponses des sujets aux trois tests intégraux.

Les analyses de Sirotnik ne révélèrent d'écart significatif d'estimation (ECIS réel -vs- ECIS simulé), ni de la moyenne, ni de la variance. Les statistiques ECIS réel s'approchaient suffisamment des paramètres de la population. Sirotnik conclut que le contexte ECIS avait peu d'effet sur l'estimation de la moyenne. Il souhaitait cependant que l'estimation de la variance soit à nouveau scrutée, car ses résultats étaient moins

évidents sur ce point. Cette étude de Sirotnik confirme surtout les résultats obtenus par Shoemaker (1970c), lesquels démontrent que le réarrangement des items d'un test ne modifie pas la performance des sujets sur ceux-ci.

Désireux de corriger certaines faiblesses de l'étude faite par Owens et Stufflebeam (1969), Impara (1972) tentait, dans un premier temps, de montrer que la technique ECIS pouvait servir efficacement à la normalisation de deux tests (l'un de vocabulaire, l'autre d'arithmétique) pour une population d'élèves dits favorisés ou défavorisés, de 4e année primaire et, en second temps, de comparer les statistiques ECIS réel de la moyenne, de la variance et du coefficient 21 de Kuder-Richardson à celles obtenues par échantillonnage unique des sujets (EUS). Souhaitant éviter l'effet d'entraînement que pouvait occasionner la présentation successive des mêmes items à un même répondant, il partagea chacune de ses deux populations en deux groupes, l'un soumis dans l'ordre aux sous-tests et au test intégral et, l'autre, en ordre inverse.

Pour les deux populations, favorisée et défavorisée, les estimateurs de la moyenne au test de vocabulaire étaient aussi précis l'un que l'autre. Toutefois, dans le cas du test d'arithmétique, l'estimation de la moyenne de la population désavantagée fut nettement plus précise que celle de l'avantagée; résultat compatible avec celui d'Owens et Stufflebeam (1969).

Les statistiques ECIS réel de l'écart-type variaient suivant la population et le test. Au test de vocabulaire, l'écart-type de la population avantagée fut mieux estimé par l'ECIS réel que celui des désavantagés; ce fut l'opposé pour le test d'arithmétique. L'estimation ECIS réel de l'écart-type manquait beaucoup plus de précision dans cette recherche que dans celles d'Owens et Stufflebeam (1969), de Plumlee (1964) et de Lord (1962), ces deux dernières étant évidemment des recherches *ex post facto*.

Bien que les résultats concernant l'estimation de la moyenne confirment ceux des autres recherches, le bilan de cette recherche laisse perplexe et nous en voulons pour preuve cet extrait du paragraphe final du rapport d'Impara:

> *"If the purpose for testing is to estimate the characteristics of the entire distribution of scores, i.e., the first four moments of the distribution, then a census test procedure should be used, unless the cost is prohibitive, then examinee sampling is recommended over matrix sampling".* (Impara, 1972, p. 64)

Novak (1974), en plus de s'intéresser à l'effet de contexte, voulait savoir si le fait d'avoir déjà répondu à des items pouvait exercer une influence sur la performance ultérieure des répondants. Il exprima trois hypothèses: 1º) la technique ECIS ne biaise pas significativement l'estimation de la moyenne et de la variance; 2º) l'exposition préalable aux items ne biaise pas significativement l'estimation ECIS de la moyenne et de la variance; 3º) à nombre égal d'observations, l'estimation ECIS de la moyenne et de la variance sera aussi précise que l'estimation EUS.

Novak expérimenta au niveau des 4e, 5e et 6e années primaires avec trois tests (vocabulaire V, épellation E et mathématique M). Il construisit des sous-tests pour chaque test, puis administra un sous-test combiné V+E+M. En moyenne, chaque sous-test combiné comprenait au total entre 18 et 24 items selon le niveau alors que, par déduction, on peut estimer que les tests intégraux en comptaient chacun entre 36 et 48.

Par son schéma expérimental, l'auteur obtint quatre classes de données: 1º) des données ECIS réel sans exposition préalable aux items; 2º) des données ECIS simulé avec exposition préalable; 3º) des données ECIS simulé sans exposition préalable; 4º) des données ECIS réel avec exposition préalable. En outre, il créa dix échantillons uniques de sujets dont le total d'observations égalait celui des estimations ECIS.

Le contexte ECIS *a priori* ne biaisa pas l'estimation de la moyenne, mais il influença significativement l'estimation de la variance quoique sans tendance systématique. L'exposition préalable aux items a provoqué un accroissement significatif de l'estimation ECIS réel de la variance. A nouveau, l'estimation ECIS réel de la moyenne s'éloigna généralement moins des paramètres que l'estimation EUS et la variance fut estimée avec la même précision par les deux techniques. Il est assez intéressant de

noter que, en dépit du fait qu'aucune tendance significative n'ait été décelée, la somme algébrique des écarts des estimations ECIS simulé à la moyenne n'est que de -0,079, tandis que celle des écarts des estimations ECIS réel égale 7,419.

Avec Feldt et Forsyth (1974), on se trouve en présence des premiers chercheurs qui identifièrent sans équivoque un certain effet de contexte sur l'estimation. Ces chercheurs entreprirent de comparer l'estimation ECIS réel de la moyenne avec l'estimation ECIS simulé.

Feldt et Forsyth expérimentèrent au niveau des 9e, 10e, 11e et 12e années scolaires avec deux tests, l'un de mathématique (36 items, 45 minutes) et l'autre en langue maternelle, l'anglais, (54 items, 35 minutes). Le test de langue maternelle fut divisé en deux sous-tests de 27 items chacun et celui de mathématique en trois de 9 items. Ils développèrent un plan expérimental où l'exposition préalable aux items fut contrôlée tout au cours de l'expérimentation.

Les analyses effectuées par Feldt et Forsyth n'indiquèrent pas de différence significative entre les moyennes au test de langue, mais, pour les quatre niveaux, les écarts de moyennes au test de mathématique se sont révélés significatifs.

Parmi les arguments invoqués par Feldt et Forsyth pour expliquer cette discordance dans leurs résultats, il y avait que les sous-tests de mathématique ne comprenaient que le quart des items du test intégral; Jacobs et Wildemann (1969), avec 10% d'items, n'observèrent aucun effet de contexte. Un autre argument s'appuie sur la nature même des matières; la langue maternelle se prête mieux à la réponse déduite, à la spéculation autant à l'intérieur d'un test court que d'un test long, tandis qu'un test à peu d'items en mathématique peut faciliter au répondant la concentration sur la résolution des problèmes; bien qu'au niveau élémentaire, Sirotnik (1970) fit usage d'un test d'arithmétique qui ne subit aucunement l'effet de contexte. En somme, le débat demeure ouvert.

Sirotnik et Wellington (1974) simplifièrent leur schéma expérimental

à l'aide de la technique ECIS en vue d'étudier l'effet de l'arrangement des items dans un test, présentation en désordre comparée à présentation ordonnée selon les contenus des items. Les contenus couverts comprenaient la mathématique, les sciences, les sciences humaines, la lecture, le langage des arts. L'étude ne révéla aucune différence appréciable sur les estimés de la moyenne, de la variance, de la variance des indices de facilité et du coefficient 20 de Kuder-Richardson.

Hill (1975) voulait créer un contexte ECIS sans effet sur l'estimation des paramètres. A une population-témoin d'élèves de 4e et 5e années, il administra sur trois demi-journées consécutives trois tests dans l'ordre, un en sciences (35 items, 45 minutes), un autre d'aptitudes (25 items, 30 minutes) et le dernier, différent du premier, également en sciences (60 items, 45 minutes). Le groupe expérimental passa un test de vingt minutes composé de 16 items tirés des plans ECIS suivants: sept sous-tests de cinq items sur le premier test de sciences, cinq sous-tests de cinq items sur le test d'aptitudes et dix sous-tests de six items sur le deuxième test de sciences.

L'estimation ECIS de la moyenne du deuxième test de sciences présentait un biais légèrement positif; sept des dix estimations surévaluaient la moyenne. Un écart significatif entre les estimations de la variance de cette dernière étude fut constaté dans le cas d'un des trois tests. Cette étude invite à nouveau à la prudence; Hill (1975) conclut d'ailleurs que l'échantillonnage combiné des items et des sujets ne peut s'employer qu'avec soin pour éviter l'effet du contexte ECIS.

BILAN

Lord (1962) affirmait que le postulat de non-influence du contexte ECIS serait acceptable en certaines occasions mais pas dans toutes. Le bilan général de la recherche sur l'effet de contexte semble, malgré tout, lui donner raison. Colligeons les indices d'un effet possible de la technique ECIS sur l'estimation: 1°) sur les quatre plans ECIS réel créés par Jacobs et Wildermann (1969), trois estimations de moyenne sur quatre sont supérieures à la moyenne de comparaison; 2°) Sirotnik (1970) exprime

quelques inquiétudes sur l'estimation ECIS réel de la variance; 3°) dans l'étude d'Impara (1972), on constate d'une part des résultats ambigus et, d'autre part, une faiblesse d'estimation ECIS de l'écart-type; 4°) à nouveau, l'estimation ECIS de la variance fait problème dans l'étude de Novak (1974) et, par ailleurs, on remarque que la somme des écarts des estimations ECIS réel à la moyenne est plus élevée que celle des écarts des estimations ECIS simulé à la moyenne; 5°) Feldt et Forsyth (1974) obtiennent une estimation biaisée de la moyenne de leur test de mathématique; 6°) Hill (1975) note aussi un écart d'estimation ECIS de la moyenne et de la variance. A cette énumération suscitant de la méfiance, on peut faire le contrepoids avec les résultats encourageants et voir ces derniers prendre le dessus. L'ambivalence du portrait d'ensemble des recherches antérieures indique, d'une part, qu'il y a encore place à d'autres études du phénomène de l'effet du contexte ECIS sur le rendement des répondants et conséquemment sur les statistiques dérivées de l'application intégrale de cette technique et, d'autre part, les centres d'intérêt de nouvelles recherches.

A part l'analyse fondée sur l'estimation des paramètres, on peut faire le bilan de la recherche portant sur l'effet du contexte ECIS en l'appréciant selon les quatre facteurs que Feldt et Forsyth (1974) qualifient de variables principales du contexte ECIS: la durée de l'épreuve, l'attitude du sujet face au test, la fatigue et le *test-wiseness* défini comme la compétence acquise en cours d'épreuve, compétence pouvant tout aussi bien se traduire au niveau de l'apprentissage du contenu qu'à celui de l'adresse à répondre aux items.

L'application intégrale de la technique ECIS suppose une réduction appréciable de la durée de l'épreuve puisqu'un sous-test ne comprend souvent qu'entre 5% et 20% du total des items dans le test intégral. Bien qu'implicite, cette réduction de la durée de l'épreuve n'a en soi fait l'objet d'aucune recherche, serait-ce simplement comme variable contrôlée. Chez Jacobs et Wildemann (1969), par les plans ECIS, on peut déduire que la durée pouvait être respectivement de 10%, 25%, 33% et 50% du temps accordé pour le test intégral. Sirotnik (1970) insère ces sous-tests dans des épreuves incluant un autre test; le contexte temporel est

donc moins définissable. Les sous-tests d'Impara (1972) demandaient entre 30% et 40% de la durée totale de chaque test intégral. Les sous-tests de Novak (1974) pouvaient prendre environ 40% du temps de passation original. Feldt et Forsyth (1974) créèrent des sous-tests de 25% et de 50% du test intégral dont le temps de passation devait être dans les mêmes proportions. Le test expérimental de Hill (1975) durait vingt minutes alors que les tests intégraux demandaient 30 ou 45 minutes. Sauf pour un cas de 10% (Jacobs et Wildemann, 1969), la durée des sous-tests est égale ou supérieure à 25% de celle du test intégral.

Quelle attitude un répondant adopte-t-il face à un test? Dans quelles conditions sera-t-il motivé à donner sa meilleure performance? Par exemple, le lien entre sa performance à un test et l'atteinte d'un objectif personnel est-il une raison qui pourrait modifier sensiblement sa performance générale à un test? Lorsqu'un test n'a pas de signification immédiate pour un répondant, quel sera son niveau de performance? Lorsqu'un test est long, aura-t-il tendance à développer un certain laxisme après avoir répondu à un certain nombre de questions? Voilà quelques questions qui taillent un peu le problème de l'attitude du répondant face au test auquel il est soumis. Vues sous cet angle, voici ce que nous révèlent les recherches recensées.

Du rapport de Jacobs et Wildemann (1969), tout ce qu'on peut extraire de renseignements est le fait que l'expérience se déroulait à l'intérieur d'un contexte de mesure du rendement d'étudiants inscrits à un cours universitaire. Sirotnik (1970) choisit des items hors test et sans lien avec l'enseignement reçu par sa population d'étudiants universitaires; il créa une situation essentiellement expérimentale qui nous permet de supposer une certaine uniformité d'attitude des sujets au cours de l'expérience. Avec Novak (1974), on se retrouve chez des élèves de 4e, 5e et 6e années auxquels on administre un test intégral faisant partie d'une batterie officielle de tests. Feldt et Forsyth (1974) inscrivent leur expérience dans une opération normale de 240 minutes d'épreuves étalées sur deux ou trois sessions; situant les sous-tests ECIS au début de la dernière session, on peut présumer que l'attitude pouvait ne pas être différente à ce moment-là de ce qu'elle était dans l'ensemble de l'opération. Hill

(1975), expérimentant avec des élèves de 4e et 5e années, administra les trois tests intégraux sur trois demi-journées consécutives pour précisément éviter l'effet de la fatigue.

En somme, on peut déduire que toutes ces expériences ont été conduites dans un contexte où chaque répondant pouvait sentir un intérêt personnel à fournir le meilleur rendement possible. Or, en normalisant un test, l'objectif poursuivi est de tracer un portrait fidèle de la performance de la population à partir de l'image réfléchie par un échantillon représentatif. Lorsque chaque sujet est ainsi sollicité, il sait que son rendement personnel n'importe aucunement, ni à lui ni à l'évaluateur. Il faut alors lui proposer des raisons altruistes pour que, malgré son désintéressement immédiat, il soit motivé à répondre le mieux possible à chaque item du test.

Plusieurs facteurs peuvent provoquer la fatigue, notamment: la difficulté du test, sa durée, son importance pour l'avenir du répondant, le climat au cours de l'épreuve. Dans l'ensemble, les chercheurs sur l'effet du contexte ECIS n'ont pas fait jouer ou n'ont pas contrôlé les facteurs inducteurs de fatigue, du moins, ils n'en font point mention.

Dans un test long, un répondant peut accidentellement développer une certaine compétence à répondre à un type particulier de questions. Après avoir répondu à plusieurs items sur un même contenu, le contenu de certaines questions peut receler des indices de réponse à d'autres; voilà en substance ce qu'entendent les chercheurs par *test-wiseness*. Jusqu'ici, les études repérées sont muettes sur ce facteur; tout ce que l'on peut dire, c'est que le contexte ECIS modifie la condition essentielle d'existence du *test-wiseness*, soit le grand nombre d'items dans le test.

Cette analyse fondée sur les facteurs constituant le contexte confirme, à notre avis, l'importance de poursuivre l'étude des effets de l'application de la technique ECIS tant sur le rendement des répondants que sur l'estimation des paramètres. Elle indique aussi des voies de recherche et suggère des hypothèses à vérifier.

AUTRES RECHERCHES

PRESENTATION

En docimologie, l'effet de contexte sur le rendement est une préoccupation constante des concepteurs de tests, d'épreuves de toute nature. A propos des particularités de la tâche de l'évalué devant un questionnaire à choix multiples, Noizet et Fabre rappelaient que:

> *"Cette tâche d'évaluation (...) est proposée dans certaines conditions de présentation séquentielle des questions, de relation entre ces questions, de représentation par le sujet des finalités de l'examen. Ces conditions, ou du moins certains éléments d'entre elles, commandent probablement les stratégies du sujet lorsque ce dernier est dans l'incertitude."*
> (Noizet et Fabre, 1975, p. 40)

On retrouve dans cette citation d'autres dimensions de l'effet du contexte ECIS. Ainsi, plusieurs chercheurs se sont penchés sur l'effet du réarrangement des items à l'intérieur d'un test ou d'un examen; citons, à titre indicatif, les travaux de Mollenkopf (1950), Brenner (1964), Sax et Cromack (1966), Shoemaker (1970c), Huck et Bowers (1972), Hambleton et Traub (1974), Barcikowski et Olsen (1975).

D'autres ont considéré le phénomène du changement des réponses (Reile et Briggs, 1966; Lynch et Smith, 1975). C'est aux Etats-Unis qu'est apparu, vers les années 60, le concept de *test-wiseness* (Millman, Bishop et Ebel, 1965), qui a suscité l'intérêt des chercheurs si l'on en juge par le nombre de rapports publiés depuis lors. D'autres chercheurs mettent en cause les caractéristiques du répondant; par exemple, Wrightsman (1962) étudia les effets du niveau d'anxiété, de la motivation et de la perception de l'importance de la tâche chez le répondant, sur son rendement à un test d'intelligence. Partant de l'hypothèse que "*...les sujets essaient de se guider sur la succession de leurs réponses pour se décider lorsqu'ils sont dans l'incertitude*" (Noizet et Fabre, 1975, p. 58), quelques docimologues, dont Flament (1963) semble le principal artisan, s'intéressent aux effets dits de *récence*.

Rapportés à la situation particulière créée par l'échantillonnage combiné des items et des sujets, tous ces facteurs ne sauraient être jugés d'intérêt égal car, si les rapports consultés éclairent la nature de chacun, leur interaction dans un contexte ECIS demeure inconnue. Dans cette dernière partie de l'étude théorique, nous allons, à l'aide de certains travaux docimologiques touchant ces problèmes plus généraux, tenter de circonscrire les principaux facteurs qui pourraient concourir à créer un effet du contexte ECIS.

ANALYSE

Le contexte se définit tant par les conditions de l'épreuve que par les caractéristiques du répondant; à titre d'exemple, on peut mentionner que le facteur durée pourra avoir une influence sur le rendement des répondants tout autant que leur degré de motivation. Sirotnik et Wellington (1974, p. 179) affirmaient que les "...*context effects include all internal (psychological and physical) and external (environmental) factors affecting examinee performance on an item-by-item basis.*" Toutefois, comme l'application intégrale de l'ECIS modifie principalement les conditions de l'épreuve, notre analyse, appuyée sur l'identification de facteurs par Feldt et Forsyth (1974, p. 74), considérera l'action de la longueur du test, de l'intérêt présenté par la tâche, de la fatigue et du *test-wiseness* sur le rendement des répondants.

La technique ECIS met le répondant en situation privilégiée pour répondre aux items contenus dans son sous-test; le nombre d'items est beaucoup moindre que dans le test intégral. Les effets de certains facteurs tels la fatigue intellectuelle, la hâte d'en finir, la compression du temps disponible, l'intérêt décroissant même et, dans certains cas, la difficulté croissante des items, peuvent conjuguer leurs effets pour altérer la performance des sujets à plusieurs items, notamment ceux apparaissant à la fin du test. On peut penser que, dans un sous-test court, les performances des répondants à une même collection d'items seront différentes de celles qu'ils auraient dans un test long. Jusqu'ici, selon nos recherches, il appert que peu ou pas de chercheurs se sont préoccupés de vérifier quelque hypothèse que ce soit dans cette direction.

Pourtant, nous croyons plausible l'hypothèse que la performance des répondants ECIS aux derniers items d'un test soit supérieure à celle des répondants au test intégral.

L'intérêt du répondant se distingue-t-il selon qu'il est en situation ECIS ou habituelle? Nous partageons l'opinion de Feldt et Forsyth (1974) qui veut que ce facteur ne s'attache guère à la situation particulière de la technique ECIS et, qu'en conséquence, on puisse considérer qu'il est sans effet notable sur le rendement des répondants. Lorsque la technique ECIS sert à l'estimation de paramètres, l'importance est mise sur les résultats de groupe et non sur la performance individuelle de chaque répondant. Dans les deux situations, on peut avoir recours au même type d'incitations aussi bien verbales qu'écrites pour stimuler l'intérêt des sujets. Ceci étant dit, nous croyons pouvoir postuler que l'intérêt des répondants pour la tâche n'est pas un facteur distinctif de la situation ECIS et, quoi qu'il en soit, qu'il ne saurait influencer significativement la performance des répondants.

Les sujets entrent en situation de *testing* avec des dispositions tant intellectuelles et affectives que physiques. Parmi ces dispositions initiales, il y en a une qui intéresse les docimologues, c'est la fatigue intellectuelle, laquelle peut se doubler d'une certaine fatigue physique. On peut présumer que le travail intellectuel ou physique demandé au répondant à un test ne peut qu'accentuer sa fatigue initiale au fur et à mesure que le temps passe, et que cette fatigue peut gêner sa performance aux derniers items. Le rendement des répondants ECIS aux derniers items d'un test pourrait être amélioré par le fait que leur tâche est considérablement réduite et que ces items sont abordés ainsi avec une plus petite dose de fatigue chez les sujets.

Le *test-wiseness* est défini par Millman, Bishop et Ebel (1965) comme: *"a subject's capacity to utilize the caracteristics and formats of the test and/or test taking situation to receive a high score"* (cité par Crehan *et al.*, 1974, p. 209). A première vue, ce facteur semblerait mieux jouer en faveur du test intégral, mais une étude de Diamond et Evans (1972) montra que le *test-wiseness* ne produisait aucun effet appréciable

sur le rendement à un test d'habiletés intellectuelles. Jusqu'ici, les chercheurs ont surtout montré que le *test-wiseness* s'apprend (Gibb, 1964; Moore, Schutz et Baker, 1966; Slakter, Koehler et Hampton, 1970) et qu'il se développe avec l'âge (Crehan, Koehler et Slakter, 1974). Donc, selon toute vraisemblance, la situation ECIS ne change pas les conditions au point qu'on puisse voir le *test-wiseness* jouer favorablement et notablement en faveur du rendement des répondants au test intégral.

Conclusion

Ainsi, que l'on se tourne du côté des recherches sur l'effet du contexte ECIS tout comme du côté des études docimologiques d'intérêt général, les données recueillies laissent place à la spéculation et appellent de nouvelles études. Plusieurs questions demeurent sans réponse; c'est à quelques-unes d'entre elles que nous voulons apporter des éléments de réponse documentés; nous les exposons au chapitre suivant.

Dans ce chapitre, nous avons tracé un bilan des recherches sur l'échantillonnage combiné des items et des sujets. Le premier volet de la recherche, les études *ex post facto*, montre la fiabilité des formules statistiques d'estimation ECIS de paramètres. Les recherches *a priori*, le second volet, ont surtout le mérite de multiplier les applications de l'ECIS. Le troisième volet d'études s'attarde spécifiquement au postulat fondamental d'absence d'effet du contexte ECIS sur la performance des répondants. Cet aspect de la recherche est complété par un aperçu des résultats de quelques travaux sur les composantes de la situation de *testing*.

CHAPITRE II

LE PLAN DE L'EXPERIENCE

Ce chapitre présente la description des deux plans expérimentaux qui fourniront des éléments de réponse aux deux questions de base que nous étudions dans cette thèse. La description de chaque plan comprend tout à la fois la structure, les postulats sous-jacents, les variables et les hypothèses ou questions spécifiques auxquelles ce plan répond plus directement.

Essentiellement le premier plan expérimental s'appuie sur la question:

Quelle influence les modifications ECIS au contexte ont-elles sur les estimations de la moyenne, de la variance et des moments d'ordres 3 et 4?

Parce que l'expérience s'appuie sur un test de 72 items duquel trois plans ECIS sont définis: le premier compte douze sous-tests de six items, le second quatre sous-tests de dix-huit items et le dernier deux sous-tests de trente-six items; chacun des paramètres sera estimé à l'aide de ces trois plans faisant ici place à trois comparaisons.

Le deuxième plan touche plus précisément le rendement des répondants et s'articule autour de la question fondamentale:

Les modifications ECIS au contexte influencent-elles différemment le rendement des répondants aux items suivant que ces derniers se trouvent au début ou à la fin du test intégral?

Encore ici, les trois mêmes plans ECIS occasionnent une triple vérification de l'interaction entre la position des items dans le test et le rendement des répondants à ceux-ci.

LES PLANS EXPERIMENTAUX

Pour établir chaque plan expérimental, nous avons tenu compte de deux aspects: ce sont d'abord la question de base, et ensuite les postulats que les recherches nous autorisent à formuler à ce stade du travail. Bien que, comme il sera dit au chapitre suivant, nous ayons intégré la collecte des données des deux plans expérimentaux dans une expérimentation unique, il apparaît plus clair de les distinguer ici car, aux niveaux du traitement et de l'analyse des données ainsi que de l'interprétation, les résultats s'analysent distinctement. Toutefois, il est évident que les deux expériences sont complémentaires et devront faire l'objet d'une appréciation conjointe en regard du but de cette étude.

LE PREMIER PLAN EXPERIMENTAL

Le premier plan expérimental a pour objet de répondre à la question de base:

> *Quelle influence les modifications ECIS au contexte ont-elles sur les estimations de la moyenne, de la variance et des moments d'ordres 3 et 4?*

Cette question appelle deux postulats que la recension des écrits nous permet de formuler:

- la technique de collecte des données par échantillonnage combiné des items et des sujets, quand elle est utilisée pour normaliser un test, change plusieurs caractéristiques du contexte dans lequel les sujets répondent

- c'est à travers leurs effets sur le rendement des répondants que ces changements fausseraient, le cas échéant, l'estimation des quatre premiers moments.

Pratiquement, notre étude porte sur la technique de collecte des données; voilà essentiellement la *variable indépendante* du premier plan

expérimental. Un traitement expérimental consistera donc à recueillir les données avec les sous-tests d'un plan d'échantillonnage combiné des items et des sujets (ECIS), alors que la comparaison se fera sur un ensemble de données obtenues par la passation du test intégral, mais distribuées ensuite par simulation du même plan d'échantillonnage combiné des items et des sujets. Comme le nombre d'items par sous-test est une dimension importante d'un plan ECIS, l'expérimentation comprendra trois plans ECIS distincts afin d'éprouver la technique du point de vue de la variation concomitante du nombre de sous-tests par plan et du nombre d'items par sous-test. A cet égard, nous avons dû nous limiter à trois plans à cause de l'envergure que prend chaque expérimentation concrète, mais nous sommes allé consciemment aux limites des combinaisons possibles pour le test retenu. Signalons que ce test [1] comprend 72 items et que les plans sont respectivement de 12 sous-tests à 6 items, de 4 sous-tests à 18 items et de 2 sous-tests de 36 items.

Les résultats des recherches, où la technique ECIS est simulée, montrent que l'estimation des paramètres est fiable. Nous avons donc décidé de contrôler la méthode de calcul des estimations et, de ce fait, de comparer les estimations calculées à partir de l'emploi réel de la technique ECIS à leurs équivalentes obtenues par simulation de la technique ECIS. Donc, la méthode de calcul des estimations est une *variable contrôlée* de cette recherche.

Les estimations des quatre premiers moments seront les *variables dépendantes* de ce plan expérimental. Pour chaque plan ECIS, les estimations tirées des données recueillies par la technique ECIS seront comparées à leurs correspondantes obtenues par simulation du plan ECIS.

Ce plan expérimental comprendra donc les opérations suivantes:

1o- identifier (choisir ou créer) un test intégral;

---

1 Nous relatons sa construction au chapitre suivant et nous le présentons dans l'appendice A.

2°- identifier une population pour fin d'inférence;

3°- constituer trois plans ECIS;

4°- recueillir des données avec les trois plans ECIS dans trois échantillons représentatifs de la population;

5°- administrer le test intégral à un autre échantillon représentatif de la population;

6°- simuler les trois plans ECIS à partir des données recueillies en 5°;

7°- calculer les estimations à partir des données des plans ECIS réel et des plans ECIS simulé;

8°- comparer les estimations tirées des plans ECIS réel à leurs correspondantes tirées des plans ECIS simulé.

Ce plan expérimental fournira des réponses aux questions spécifiques suivantes:

. Etant donné un contexte de normalisation d'un test de 72 items à une population cible:

.. pour un plan ECIS à 12 sous-tests de 6 items par sous-test

Question 1

L'estimation de la moyenne obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Question 2

L'estimation de la variance obtenue en appliquant la technique ECIS

à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Question 3

L'estimation du moment d'ordre 3 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Question 4

L'estimation du moment d'ordre 4 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

.. pour un plan ECIS à 4 sous-tests de 18 items

Question 5

L'estimation de la moyenne obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Question 6

L'estimation de la variance obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Question 7

L'estimation du moment d'ordre 3 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Question 8

L'estimation du moment d'ordre 4 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

.. pour un plan ECIS à 2 sous-tests de 36 items par sous-test

Question 9

L'estimation de la moyenne obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Question 10

L'estimation de la variance obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Question 11

L'estimation du moment d'ordre 3 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Question 12

L'estimation du moment d'ordre 4 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

LE DEUXIEME PLAN EXPERIMENTAL

Le deuxième plan expérimental répondra à un ensemble de questions spécifiques qui se dégagent de la deuxième question de base:

Les modifications ECIS au contexte influencent-elles différemment le rendement des répondants aux items suivant que ces derniers se trouvent au début ou à la fin du test intégral?

Le fait de distribuer les items du test intégral dans plusieurs sous-tests entraîne des modifications du contexte de présentation des items à trois niveaux. D'abord, la séquence d'items se trouve changée en ce sens qu'un même item, en passant du test intégral à un sous-test, se voit encadré par de nouveaux items; ensuite, le rang d'un item passant du test intégral à un sous-test peut aussi changer et même de beaucoup; enfin, le nombre d'items dans la séquence est réduit.

Le sommaire des recherches portant sur le réarrangement des items à l'intérieur d'un test révèle que l'effet de ce réarrangement sur le rendement des répondants, dans des conditions normales, serait négligeable. Barcikowski et Olsen (1975), Klosner et Gellman (1973, Huck et Bowers (1972), Shoemaker (1970c), Sax et Cromack (1966), et, Brenner (1964), comme conclusion à leurs travaux, opinent dans le sens d'un effet négligeable du réarrangement sur la performance des répondants. Par contre, Hambleton et Traub (1974) concluent sans ambages à un effet significatif du réarrangement alors que Whitely et Davis (1976) supposent que les items peuvent avoir une influence les uns sur les autres dans certaines conditions, comme lorsque la capacité des sujets à répondre aux items est faible. En conséquence, nous avançons l'hypothèse que si la technique ECIS devait influencer le rendement des répondants, ce ne serait pas surtout dû au réarrangement des items mais plutôt à la réduction du nombre d'items précédant tel autre item donné. Et c'est cette hypothèse intermédiaire que ce plan expérimental vise à confirmer.

Lorsqu'on applique la technique ECIS, en se retrouvant dans les sous-tests, les derniers items du test se voient précédés d'un plus petit nombre d'items. Par contre, on ne peut en dire autant des premiers items du test qui, eux, ne se trouvent pas déjà précédés de plusieurs autres items. La *variable indépendante* de ce second plan expérimental sera donc la position des items dans le test, soit le lieu où un groupe d'items se retrouve dans

la séquence de présentation des items du test intégral. L'étude se restreindra à comparer essentiellement les items du premier quart d'un test (les 18 premiers items) à ceux du quatrième (les 18 derniers), soit les deux états de la variable indépendante.

Le rendement des répondants, la *variable dépendante*, sera mesuré par les indices de facilité des items, ceux-ci étant les rapports entre le nombre de bonnes réponses à un item et le nombre de répondants. Nous faisons le raisonnement que les indices de facilité étant le reflet de la performance des sujets aux items, ces indices varieront en fonction de la difficulté qu'éprouvent les répondants à répondre aux items. Puisque nous modifions les conditions de passation en passant du test aux sous-tests, nous pouvons postuler que des conditions différentes de contexte peuvent entraîner des performances différentes de la part des répondants provenant d'échantillons représentatifs d'une même population.

Plusieurs variables feront l'objet de contrôle: notamment le nombre d'items dans chaque quart, la moyenne et la variance de leurs indices de facilité, dans les premier et dernier quarts du test intégral, de manière à assurer leur comparabilité. Les trois plans ECIS utilisés permettront de faire varier sous contrôle la longueur (6,18,36) des séquences d'items.

Enfin, ce plan expérimental comprendra les mêmes cinq premières étapes que le précédent, soit

1°- identifier (choisir ou créer) un test intégral;

2°- identifier une population aux fins d'inférence;

3°- constituer trois plans ECIS;

4°- recueillir des données avec les trois plans ECIS dans trois échantillons représentatifs de la population;

5°- administrer le test intégral à un autre échantillon représentatif de la population;

suivies de

6°- calculer les indices de facilité des items à partir des réponses au test intégral et aux trois plans ECIS réel;

7°- vérifier la comparabilité des deux quarts du test intégral dans chaque cas;

8°- comparer les quatre séries d'indices de facilité pour les 18 premiers items et pour les 18 derniers items.

Ce plan expérimental fournira des réponses aux questions spécifiques suivantes:

. Etant donné un contexte de normalisation d'un test de 72 items à une population cible:

.. où un plan ECIS à 12 sous-tests de six items est employé pour recueillir les données

Question 1

Aux 18 premiers items du test intégral, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

Question 2

Aux 18 derniers items du test intégral, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

.. où un plan ECIS à quatre sous-tests de 18 items est employé pour recueillir les données

Question 3

Aux 18 premiers items du test intégral, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

Question 4

Aux 18 derniers items du test intégral, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

.. où un plan ECIS à deux sous-tests de 36 items est employé pour recueillir les données

Question 5

Aux 18 premiers items du test intégral, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

Question 6

Aux 18 derniers items du test intégral, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

SOMMAIRE

Nous venons de décrire les principales composantes des deux plans expérimentaux, corps de cette recherche. En particulier, les postulats, les variables et les questions spécifiques y sont explicités. Les étapes de chaque plan sont énumérées et serviront de structure dans le prochain chapitre décrivant l'expérience, sa préparation et son exécution.

CHAPITRE III

L'EXPERIENCE

Le déroulement de l'expérience décrit la réalisation des cinq étapes communes aux deux plans expérimentaux. En premier lieu, figure la construction du test, suivie de l'identification de la population. Viennent ensuite la détermination des plans ECIS et la composition des échantillons. Le chapitre se termine par la description des étapes de simulation des plans ECIS et la relation de l'expérimentation proprement dite.

LE TEST

L'analyse de la situation révéla très tôt qu'il serait plus simple de construire un test *ad hoc* avec des items empruntés à un test existant déjà et normalisé pour une population comparable. Bhushan (1974) avait traduit, adapté et normalisé pour les classes de 5e à 9e année de langue française du Québec les versions EM et FM du test *Otis Quick-Scoring Mental Ability Test, Beta Test* qui comptaient au total 160 items à choix multiple (4 ou 5) et à réponse correcte unique. Le travail de normalisation de ce test procurant une estimation des indices de facilité de ces items pour une population-cible convenable aux fins de la présente étude, nous avons opté pour leur emploi comme réservoir premier duquel seraient extraits les items composant un test adapté à nos besoins particuliers.

En procédant ainsi, nous obtenions un test de puissance, avec un nombre suffisant d'items arrangés pour satisfaire aux exigences des deux plans expérimentaux.

Voici donc les étapes de construction du test:

1°- De la banque initiale de 160 items furent retranchés ceux dont l'indice de facilité était inférieur à ,100 ou supérieur à ,899; il en est resté 131.

2°- Pour construire un test à quatre parties égales et comparables, les 131 items furent d'abord couplés en fonction de l'indice de facilité et du contenu de chacun, réduisant ainsi le bassin à 36 paires d'items. Par la suite, chaque paire d'items fut attribuée au hasard à l'une ou l'autre des deux combinaisons de parties du test, premier et quatrième quarts ou deuxième et troisième quarts. Finalement, un des deux items couplés était tiré au sort comme item du premier quart ou du deuxième selon la combinaison à laquelle la paire était destinée; ainsi, nous avions les 72 items répartis dans quatre sous-ensembles de 18 items, les quarts du test.

3°- Ayant besoin surtout de comparer les premier et quatrième quarts entre eux, nous avons décidé d'ordonner aléatoirement les items dans chacun des quarts de sorte qu'*a priori*, seuls ces derniers sont comparables. Le test intégral et les sous-tests sont présentés à l'appendice A.

LA POPULATION

Uniquement en raison de leur accessibilité, la population a d'abord été définie par les quatre commissions scolaires régionales suivantes: du Saguenay, du Bas-Saint-Laurent, de la Matapédia et de la Baie-des-Chaleurs. Elle fut, par la suite, limitée aux élèves de secondaire I pour deux raisons utilitaires: le test intégral était de difficulté moyenne pour eux, ce qui nous semblait préférable; ces élèves pouvaient en grande majorité répondre aux 72 questions du test en moins de quarante-cinq (45)

minutes, soit la plus courte période d'enseignement continu, laquelle correspond à la durée de passation.

LES ECHANTILLONS DE SUJETS

Au moment de déterminer les échantillons de sujets, quelques considérations s'imposèrent: d'abord, pour que les plans ECIS puissent être comparables, la proportion de sujets tirés de chacune des commissions scolaires devait être la même d'un échantillon à l'autre et aussi, le nombre d'observations devait être constant d'un plan ECIS à l'autre. Au moment de planifier, nous avions retenu le modèle de répartition présenté au tableau 2. Un échantillon de sujets devaient répondre au test intégral en vue de simuler ultérieurement les mêmes plans ECIS.

Une fois l'expérimentation terminée, la répartition subit quelques modifications concernant le total de sujets par plan ECIS, mais la proportion des sujets entre les commissions scolaires est demeurée sensiblement équivalente d'un plan ECIS à l'autre. Le tableau 3 résume le produit final. Les 171 répondants au test intégral servirent à simuler les trois mêmes plans ECIS.

TABLEAU 2

Répartition hypothétique des sujets par commission scolaire et par plan ECIS

Plan ECIS	(1) Groupe-classe par régionale	(2) Nombre de sujets par groupe-classe	(3) Total de sujets (1) x (2) x 4
(2,36)	1	15	60
(4,18)	1	30	120
(12,6)	3	30	360
Test intégral	2	30	240

TABLEAU 3

Nombre de groupes-classes, nombre total de sujets et nombre d'observations par plan ECIS et au test intégral, tirés de l'expérimentation

Plan ECIS	Nombre de groupes-classes	Nombre de sujets	Nombre d'observations
(2,36)	4	65	2340
(4,18)	5	134	2412
(12,6)	14	373	2238
Test intégral	15	171	ne s'applique pas

Les plans ECIS

Les plans ECIS se distinguent par leurs trois dimensions: le nombre t de sous-tests, le nombre k d'items dans chaque sous-test et le nombre n de répondants à chaque sous-test; d'où l'habitude de présenter un plan ECIS par le triplet (t, k, n). La troisième valeur du triplet avait sa raison d'être tant que l'on exigeait que le nombre de répondants à chaque sous-test soit le même pour tous; or, les nouveaux algorithmes d'estimation des paramètres la rendent possible même si le nombre de répondants à chaque sous-test varie. Nous désignerons donc chaque plan par le doublet (t, k).

Les critères suivants ont guidé le choix des plans ECIS:

. Le produit t k = K, où t est le nombre de sous-tests, k le nombre d'items par sous-test et K le nombre d'items dans le test intégral.

. En supposant une distribution normale, ce serait l'augmentation du nombre d'items par sous-test qui réduirait davantage l'erreur-type d'estimation.

. La variation de la longueur des sous-tests devait être au moins le double entre deux plans pour que l'écart entre chaque plan se caractérise par une modification sensible du contexte.

Le test intégral comprenant 72 items fut donc divisé en deux sous-tests de 36 items, en quatre sous-tests de 18 items et en 12 sous-tests de six items pour obtenir les plans ECIS (2,36), (4,18) et (12,6). Les deux sous-tests du plan (2,36) étaient constitués comme suit: le premier des 36 premiers items du test intégral, le second des 36 derniers. De façon analogue, les 18 premiers items du test intégral formaient le premier sous-test du plan (4,18), les 18 suivants le deuxième et ainsi de suite pour les deux derniers sous-tests. Chaque tranche de six items pris dans l'ordre, du premier au soixante-douzième, déterminait chacun des douze sous-tests du plan (12,6). Le tableau 4 montre la subdivision du test intégral en sous-tests pour chacun des plans ECIS.

LA SIMULATION DES PLANS ECIS

Le premier plan expérimental prévoit une comparaison d'estimations de paramètres issues, d'une part, de l'application d'un plan ECIS et, d'autre part, d'une simulation de ce même plan sur des données recueillies par administration du test intégral.

Pour simuler les plans ECIS, nous avons constitué un échantillon de répondants au test intégral dont la répartition [illegible] les commissions scolaires respectait celle établie pour [illegible] ECIS. Dans le cas des plans (2,36) et (4,18), un même répondant n'avait à constituer qu'un seul sous-échantillon de sujets à un sous-test; par contre, pour le plan (12,6), certains répondants feraient partie de trois sous-échantillons; évidemment, chaque sous-échantillon de sujets étant associé à un sous-test distinct, ce sont des réponses à des items différents qui serviraient dans chaque cas.

En déterminant les sous-échantillons de sujets pour chaque plan ECIS, nous avons respecté aussi fidèlement que possible les règles suivantes:

TABLEAU 4

Subdivision du test intégral en sous-tests pour chacun des plans ECIS

Plan ECIS	Items	Numérotation des sous-tests
(2,36)	1 à 36	1
	37 à 72	2
(4,18)	1 à 18	1
	19 à 36	2
	37 à 54	3
	55 à 72	4
(12,6)	1 à 6	1
	7 à 12	2
	13 à 18	3
	19 à 24	4
	25 à 30	5
	31 à 36	6
	37 à 42	7
	43 à 48	8
	49 à 54	9
	55 à 60	10
	61 à 66	11
	67 à 72	12

. Aucun répondant ne peut être choisi plus d'une fois dans un même sous-échantillon;

. Aucun répondant choisi dans les sous-échantillons des sous-tests du premier quart du test intégral ne peut être choisi comme sujet dans les sous-échantillons des sous-tests du dernier quart;

. La sélection des répondants doit respecter la représentation proportionnelle des commissions scolaires telle qu'elle a été établie lors de l'application des plans ECIS à la collecte des données.

Au terme de l'expérimentation, 171 sujets, provenant en nombre à peu près égal des commissions scolaires, avaient répondu au test intégral; il s'agissait là du réservoir pour la simulation des trois plans ECIS. Pour respecter uniformément les règles, nous avons adopté dans les trois cas une démarche unique de simulation; elle comprenait les étapes suivantes:

LA CONSTRUCTION DE DEUX ECHANTILLONS REPRESENTATIFS

Les 171 répondants sont partagés en deux échantillons, l'un de 86 répondants, désigné pour la construction des sous-échantillons aux sous-tests du premier, puis du second quart du test intégral; l'autre de 85 répondants, désigné pour la construction des sous-échantillons aux sous-tests du quatrième, puis du troisième quart. Nous avons construit les deux échantillons par prélèvement systématique des répondants, les impairs au premier échantillon, les pairs au second, respectant ainsi la représentativité des commissions scolaires dans les deux échantillons.

LA CONSTRUCTION DES SOUS-ECHANTILLONS RELATIFS AUX SOUS-TESTS

Pour chaque plan ECIS, nous avons d'abord construit les sous-échantillons relatifs aux sous-tests des premier et quatrième quarts à partir de leur échantillon désigné respectif. Pour le plan (2,36), le sous-échantillon de 37 répondants au premier sous-test a été construit en prenant les répondants pairs, soit 43; puis, en retirant systématiquement les septièmes,

il en est resté 37. Le deuxième sous-test demandait un sous-échantillon de 28 répondants construit en conservant tous les troisièmes sujets de l'échantillon. Le plan (4,18) a été simulé de la même façon, les sous-échantillons aux premier et quatrième sous-tests construits d'abord, ensuite ceux des deuxième et troisième. Au moment de simuler le plan (12,6), il a fallu utiliser de nouveau les deux échantillons deux fois au complet et une troisième fois de façon partielle; les sous-échantillons aux six premiers et aux six derniers sous-tests furent constitués d'abord. Nous avons encore une fois procédé par échantillonnage systématique. Le tableau 5 présente le sommaire de l'opération globale d'application et de simulation des trois plans ECIS.

L'EXPERIMENTATION

Une fois le test et les sous-tests construits, mis en page et imprimés, la mise sous enveloppes engagea l'expérimentation proprement dite. Dans le cas de chaque plan ECIS, les sous-tests furent empilés systématiquement, mais le premier sous-test empilé était différent d'une enveloppe à l'autre. Par exemple, pour le plan (4,18), dans une enveloppe les sous-tests étaient empilés dans l'ordre 1er, 2e, 3e, 4e, 1er, 2e, 3e, 4e, etc., alors que dans la suivante, ils étaient disposés dans l'ordre 2e, 3e, 4e, 1er, 2e, 3e, 4e, 1er, etc., et ainsi de suite pour les enveloppes suivantes.

Nous n'avons choisi ni les écoles ni les professeurs-expérimentateurs; c'est le consentement des professeurs à collaborer qui détermina les deux à la fois. Comme, dans les polyvalentes, les groupes-classes se forment au gré des matières et des niveaux, nous avons résolu de n'utiliser les services que d'un seul professeur par école, ceci pour éviter qu'un élève soit sujet de plusieurs échantillons. Dans toutes les écoles, au moins les sous-tests d'un plan ECIS et le test purent être administrés. Une fois un professeur identifié, une enveloppe était attribuée aléatoirement à chacun de ses groupes-classes, étant entendu qu'un groupe recevait le test. Les instructions lui étaient ensuite transmises par écrit (appendice B), lues et commentées.

Etant donné les horaires de chacun des professeurs, il ne fut pas

TABLEAU 5

Le nombre d'items, le nombre de répondants et le nombre d'observations pour chaque plan ECIS et chaque sous-test, application réelle ou simulée

Plan ECIS	No du sous-test	Nombre d'items	Nombre de sujets		Nombre d'observations	
			Réel	Simulé	Réel	Simulé
(2,36)	1	36	37	37	1332	1332
	2	36	28	28	1008	1008
Total		72	65	65	2340	2340
(4,18)	1	18	30	31	540	558
	2	18	33	33	594	594
	3	18	35	35	630	630
	4	18	36	36	648	648
Total		72	134	135	2412	2530
(12,6)	1	6	32	32	192	192
	2	6	32	32	192	192
	3	6	33	33	198	198
	4	6	30	30	180	180
	5	6	31	31	186	186
	6	6	31	31	186	186
	7	6	31	31	186	186
	8	6	29	29	174	174
	9	6	27	27	162	162
	10	6	30	30	180	180
	11	6	34	34	204	204
	12	6	33	33	198	198
Total		72	373	373	2238	2238

possible d'aménager une heure de passation unique, mais nous avons une indication au tableau 6 de la répartition des périodes d'expérimentation. La période d'expérimentation s'étendit de la mi-mai 1978 à la première semaine de juin. A l'intérieur d'une école, nous avions demandé aux professeurs d'effectuer autant que possible l'expérimentation dans la même demi-journée, mais cela fut impraticable pour quelques-uns. Cependant, tous les professeurs ont administré le test et les sous-tests à leurs groupes respectifs au cours d'une même journée.

Dans ce chapitre, nous avons rapporté la façon dont nous avons préparé l'expérimentation, les moyens mis en oeuvre pour assurer l'atteinte des objectifs de la recherche et réduire au maximum les possibilités d'erreurs expérimentales, les diverses opérations qui composaient l'étape pratique de la recherche. Au chapitre suivant, nous dégageons les résultats de cette expérimentation.

TABLEAU 6

Fréquence des groupes-classes par période d'administration du test et des sous-tests

Période	Nombre de groupes-classes
8h30 à 9h29	4
9h30 à 10h29	3
10h30 à 11h29	3
13h00 à 13h59	6
14h00 à 14h59	5
15h00 à 15h59	1
total	25

CHAPITRE IV

LES RESULTATS

Ce chapitre respecte la partition en deux plans expérimentaux et l'ordre des questions spécifiques de ces derniers. Pour le premier plan, nous exposons les résultats généraux, tout particulièrement les estimateurs, puis le résultat des analyses propres à chaque question; nous terminons cette partie par une discussion générale sur chaque moment estimé et sur la question de base et, finalement, une mise en relation de nos résultats avec ceux des autres chercheurs. Quant au deuxième plan, nous en présentons les données brutes, puis le résultat de l'analyse propre à chaque question; le chapitre est complété par une discussion générale unique sur l'ensemble des résultats obtenus pour ce plan.

En guise de réponse à la première question de base, notre étude fournit une autre preuve empirique qu'en appliquant la technique d'échantillonnage combiné des items et des sujets à la collecte des données aussi bien qu'au calcul des estimateurs, 1°) que les estimateurs de la moyenne peuvent différer significativement de leurs correspondants tirés de simulations de plans ECIS, 2°) que les estimateurs de la variance et du moment d'ordre 3 ne semblent pas ou sont peu influencés, et 3°) que l'estimation du moment d'ordre 4 puisse être biaisée.

Les faits observés dans la deuxième expérimentation indiquent que la position des items dans le test intégral n'est pas un facteur significatif de variation de la performance des sujets lorsque ceux-ci répondent aux items dans les sous-tests d'un plan ECIS. S'il existe une influence du contexte ECIS sur la performance des répondants, cette influence se manifeste plutôt uniformément sur l'ensemble des items composant le test.

LE PREMIER PLAN EXPERIMENTAL

Ce plan mettait en comparaison les estimations des quatre premiers moments d'un test tirées de différents plans ECIS appliqués intégralement ou simulés. Le tableau 7 présente les estimations [1] des quatre premiers moments et leurs erreurs-types pour chacun des plans ECIS, réels ou simulés.

QUESTION 1

Pour le plan ECIS (12,6), l'estimation de la moyenne obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en la simulant?

La question posée équivaut à comparer les estimations ECIS de la moyenne de deux échantillons de la population, l'un ayant répondu à des sous-tests ECIS, l'autre ayant répondu au test intégral. Pour vérifier l'hypothèse que l'écart entre les moyennes est négligeable, nous avons effectué une analyse de variance dont le résumé figure au tableau 8.

Le résultat de cette analyse nous laisse donc formuler l'hypothèse que les modifications au contexte provoquées par le plan ECIS (12,6) ont un effet sur l'estimation de la moyenne. La moyenne de l'échantillon de répondants aux sous-tests à six items est inférieure à celle des répondants au test intégral (voir tableau 7) indiquant que cet effet du contexte ECIS, dans le cas d'un plan (12,6), irait dans le sens d'une diminution du rendement des premiers.

---

1 Les estimations ECIS et leurs erreurs-types ont toutes été calculées à l'aide du programme informatique intitulé *Estimation of Parameters of Universe Scores through Multiple Matrix Sampling*, mis au point et revisé la dernière fois en juin 1977 par le docteur David Shoemaker qui l'a gracieusement mis à notre disposition.

TABLEAU 7

Les estimations de paramètres et leurs erreurs-types en fonction des plans ECIS (réel ou simulé)

Plan ECIS	TYPE	# sous-tests	Taille sous-test	# observations	$\bar{X}$	$S_{\bar{X}}$	S^2	S_{S2}	M_3	S_{M_3}	M_4	S_{M_4}
(2,36)	R	2	36	2340	37,705	2,133	154,446	19,94	-399,	848,	51861,	3115,
	S	2	36	2340	38,258	0,315	102,892	10,52	567,	492,	24507,	7532,
(4,18)	R	4	18	2512	36,087	0,813	91,178	21,08	323,	305,	12194,	5997,
	S	4	18	2530	39,986	1,423	94,535	19,76	- 76,	282,	16355,	4939,
(12,6)	R	12	6	2238	34,348	1,804	110,099	22,84	-502,	628,	15650,	15105,
	S	12	6	2238	38,659	2,260	104,588	20,61	606,	361,	47147,	26007,

LEGENDE

- # nombre
- $\bar{X}$ moyenne
- $S_{\bar{X}}$ erreur-type de la moyenne
- S^2 variance
- S_{S2} erreur-type de la variance
- M_3 moment d'ordre 3
- S_{M_3} erreur-type du moment d'ordre 3
- M_4 moment d'ordre 4
- S_{M_4} erreur-type du moment d'ordre 4
- R réel
- S simulé

TABLEAU 8

Analyse de la variance entre les moyennes de l'échantillon de sujets au plan ECIS (12,6) réel et de l'échantillon de sujets au plan ECIS (12,6) simulé.

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	3466,0574	1	3466,0574	32,5736 *
ERREUR	79166,7500	744	106,4069	
TOTAL	82632,7500	745		

* Significatif à $p \leq 0,01$

QUESTION 2

Pour le plan ECIS (12,6), l'estimation de la variance obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Les estimateurs obtenus sont-ils tous deux estimateurs de la variance d'une seule population? Nous avons vérifié cette question à l'aide du test F de R.-A. Fisher: le rapport calculé étant de 1,053 alors que la valeur requise pour que les variances soient considérées différentes est $F_{,99}(372,372)=1,28$, on peut maintenir l'hypothèse que les deux variances estiment la variance d'une même population.

QUESTION 3

Pour le plan ECIS (12,6), l'estimation du moment d'ordre 3 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Pour établir la probabilité que $M'_3 - M''_3 = 0$, nous avons fait appel au test $z = \frac{M'_3 - M''_3}{\sqrt{S_{M'_3} + S_{M''_3}}}$ (Guilford, formules 9.1 et 9.2, p. 174-175). Nous obtenons un $z = 1,53 < 1,96$ ($p < 0,05$), donc non significatif. D'où le maintien de l'hypothèse nulle qui indique que, selon toute vraisemblance, ces deux statistiques estiment le paramètre μ_3 d'une même population.

QUESTION 4

Pour le plan ECIS (12,6), l'estimation du moment d'ordre 4 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Pour répondre à la question 4, nous avons encore comparé les deux estimateurs par le même test z qu'à la question 3. Un $z = 1,047 < 1,96$ ($p < 0,05$) n'infirme pas l'hypothèse nulle. Il y a donc lieu de croire que les deux statistiques obtenues estiment le moment d'ordre 4 d'une même population.

QUESTION 5

Pour un plan ECIS (4,18), l'estimation de la moyenne obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en la simulant?

L'analyse de la variance (tableau 9) infirme l'hypothèse nulle; il semble donc que ces deux statistiques de la moyenne proviennent d'échantillons de sujets distincts. Le fait d'avoir recueilli les données avec la technique ECIS aurait donc modifié sensiblement les résultats pour que l'estimation de la moyenne soit différente de celle que donne la technique ECIS simulée.

QUESTION 6

Pour le plan ECIS (4,18), l'estimation de la variance obtenue

TABLEAU 9

Analyse de la variance entre les moyennes de l'échantillon de sujets au plan ECIS (4,18) réel et de l'échantillon de sujets au plan ECIS (4,18) simulé.

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	1022,3345	1	1022,3345	11,0098 *
ERREUR	24792,7656	267	92,8568	
TOTAL	25815,0977	268		

* Significatif à $p \leq 0,01$

en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Le rapport F entre les deux variances, étant égal à 1,04, confirme la probabilité à 99% qu'elles soient homogènes, car $F_{,99}$ (134,133) = 1,50. La collecte des données par la technique ECIS n'aurait donc pas sensiblement influencé l'estimation de la variance.

QUESTION 7

Pour le plan ECIS (4,18), l'estimation du moment d'ordre 3 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Dans ce cas-ci, le $z = 0,96 < 1,96$ ($p \leq 0,05$) n'infirme pas l'hypothèse nulle. On peut donc conclure que ces deux estimateurs se rapportent au moment d'ordre 3 d'une même population.

QUESTION 8

Pour le plan ECIS (4,18), l'estimation du moment d'ordre 4 obtenue en appliquant la technique ECIS à la collecte des

données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

A cette question, nous donnerons le même modèle de réponse qu'à la question 4. A 0,536, le z obtenu, de toute évidence, soutient l'hypothèse nulle; en effet, il eût fallu un z = 1,96 pour la rejeter avec une probabilité 0,05 de se tromper. Pour le plan ECIS (4,18), l'estimateur du moment d'ordre 4 obtenu en appliquant la technique ECIS à la collecte des données est équivalent à celui obtenu en simulant la technique ECIS.

QUESTION 9

Pour le plan ECIS (2,36), l'estimation de la moyenne obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en la simulant?

L'analyse de la variance (tableau 10) révèle une différence non significative entre les statistiques de la moyenne ECIS réel et ECIS simulé. Par conséquent, l'hypothèse que les deux statistiques estiment le paramètre μ d'une même population est retenue.

TABLEAU 10

Analyse de la variance entre les moyennes de l'échantillon de sujets au plan ECIS (2,36) réel et de l'échantillon de sujets au plan ECIS (2,36) simulé.

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	9,9748	1	9,9748	0,0775
ERREUR	16470,7734	128	128,6779	
TOTAL	16480,7461	129		

QUESTION 10

Pour le plan ECIS (2,36), l'estimation de la variance obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Le rapport F = 1,50 entre les deux variances indique qu'elles sont homogènes puisque $F_{,99}$ (64,64) = 1,81. D'où l'implication que la collecte des données par la technique ECIS n'a pas modifié sensiblement les résultats.

QUESTION 11

Pour le plan ECIS (2,36), l'estimation du moment d'ordre 3 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Le test z, à 1,006, n'occasionne pas le rejet de l'hypothèse nulle à $p < 0,05$. Il semble donc que pour le plan ECIS (2,36), l'estimateur du moment d'ordre 3 obtenu en appliquant la technique ECIS à la collecte des données est équivalent à celui obtenu en simulant la technique ECIS.

QUESTION 12

Pour le plan ECIS (2,36), l'estimation du moment d'ordre 4 obtenue en appliquant la technique ECIS à la collecte des données est-elle statistiquement équivalente à celle obtenue en simulant la technique ECIS?

Un z = 3,355, donc significatif à $p < 0,05$ permet le rejet de l'hypothèse nulle. D'où la plausibilité de l'affirmation que pour le plan ECIS (2,36), l'estimateur du moment d'ordre 4 obtenu en appliquant la technique ECIS à la collecte des données est statistiquement différent de celui obtenu en simulant la technique ECIS.

DISCUSSION

Nous discuterons successivement des résultats du premier plan expérimental par rapport à chacun des moments, pour ensuite proposer une réponse globale à la question de base du premier plan expérimental. Nous terminerons avec quelques considérations relatives à la comparaison de nos résultats avec ceux des recherches antérieures.

La moyenne

Les plans ECIS (12,6) et (4,18) réel ont tous deux sous-estimé la moyenne si l'on en juge par comparaison avec les estimations émanant des plans ECIS simulé. Comme la technique ECIS a servi à la collecte des données dans un cas et pas dans l'autre, il semble que les modifications au contexte occasionnées par les deux plans soient la cause de cette différence significative entre les estimateurs. Par contre, les estimations ECIS (2,36) réel et simulé sont comparables, laissant ainsi supposer que l'on puisse appliquer un plan ECIS à la collecte des données sans entraîner des perturbations sensibles dans l'estimation de la moyenne; il suffit que les sous-tests constituent pour les répondants une tâche d'envergure stimulante, c'est-à-dire représentant un certain défi pour les répondants. Nous y reviendrons.

La variance

De toute évidence, les modifications ECIS apportées au contexte n'ont pas influencé l'estimation de la variance suffisamment pour que nos tests notent une différence significative entre les estimateurs ECIS réel et ECIS simulé de chaque plan ECIS. Il est cependant regrettable de ne pouvoir juger de la variation des estimations d'un plan à l'autre. Mais on peut expliquer la sur-estimation de la variance issue du plan ECIS réel (2,36) par une erreur d'échantillonnage, car aucune autre raison ne la justifie.

Le moment d'ordre 3

Au même titre que pour la variance, on peut raisonnablement penser que l'estimation du moment d'ordre 3 est peu affectée par les conditions particulières ECIS de collecte des données. Il apparaît que la statistique du coefficient d'asymétrie tirée des estimateurs de M_2 et M_3 sera un estimateur efficace. Le tableau 11 présente les coefficients d'asymétrie calculés avec la formule $A_3 = \frac{M_3}{\sqrt{M_2^3}}$ (Peatman, 1963, p. 72-77). On remarque que la variation des valeurs, tant pour les cas d'application de la technique ECIS que pour ceux de simulation, oscille autour de 0, laissant supposer que la population parente aurait, à peu de choses près, une distribution normale.

TABLEAU 11

Les coefficients d'asymétrie A_3 des distributions des échantillons aux plans ECIS réel ou simulé

TYPE DE PLAN	Plan		
	(12,6)	(4,18)	(2,36)
REEL	-0,434	+0,371	-0,208
SIMULE	+0,567	-0,083	+0,563

Le moment d'ordre 4

Les plans ECIS réel (12,6) et (4,18) donnent des estimateurs comparables dans chaque cas. Comme pour la variance, le plan (2,36) réel paraît surévaluer le moment d'ordre 4; peut-on soutenir l'hypothèse d'une erreur d'échantillonnage? Partant du point de vue que le moment d'ordre 4 nous intéresse surtout parce qu'il sert à calculer le coefficient d'aplatissement avec la formule $A_4 = \frac{M_4}{M_2^2}$ (Peatman, 1963, p. 72-77), nous avons effectivement calculé les coefficients d'aplatissement des échantillons (voir tableau 12). Or, sur six coefficients, cinq marquent une très légère

déconcentration des fréquences, le sixième (plan (12,6) simulé) supposant plutôt une tendance à la concentration. Somme toute, on serait finalement tenté de dire qu'il s'agit d'estimateurs d'une distribution normale légèrement aplatie et, ce faisant, que les estimations de M_4 ne sont peut-être pas aussi différentes qu'il n'y paraît. On peut remarquer que pour le plan ECIS (12,6), les deux coefficients diffèrent peu et qu'ils manifestent tous deux une faible tendance à la déconcentration. En outre, dans le cas du plan ECIS (2,36), l'écart significatif constaté entre les deux estimateurs perd quelque peu de son importance et peut être attribué à cette même erreur d'échantillonnage dont nous faisons état pour la variance. En dépit de cet écart, nous croyons donc que la technique ECIS de collecte des données conduirait à des estimateurs fiables du moment d'ordre 4 dans la plupart des situations comparables aux nôtres.

TABLEAU 12

Les coefficients d'aplatissement A_4 des distributions des échantillons aux plans ECTS réel ou simulé

TYPE DE PLAN	Plan		
	(12,6)	(4,18)	(2,36)
REEL	2,174	1,467	1,289
SIMULE	2,315	1,830	4,310 *

* Le seul coefficient marquant une tendance à la concentration.

La question de base

Comme réponse à la première question de base, le résultat de la présente recherche, compte tenu de ses limites, constitue une preuve empirique que lorsque la technique d'échantillonnage combiné des items et des sujets est appliquée à la collecte des données aussi bien qu'au calcul des estimateurs,

1°) les estimateurs de la moyenne peuvent différer significativement de leurs correspondants tirés de simulations de plans ECIS;

2°) les estimateurs de la variance et du moment d'ordre 3 ne semblent pas ou peu influencés, constatation fondée sur leur comparaison avec leurs correspondants obtenus par simulation de la technique ECIS.

Il semble que la collecte des données par un plan ECIS puisse biaiser l'estimation du moment d'ordre 4. Mais on peut raisonnablement penser qu'en prenant un nombre suffisant de sous-tests, il est possible d'obvier à cette difficulté.

Les autres recherches

Les recherches sur l'effet de contexte ont montré, jusqu'ici, que la collecte des données par la technique ECIS n'influence guère l'estimation de la moyenne, mais que dans certaines circonstances (Feldt et Forsyth, 1974; Hill, 1975), on pouvait constater un écart significatif et toujours une tendance à la sur-estimation. Nous observons une sous-estimation de la moyenne pour les plans ECIS (12,6) et (4,18); comment cela s'explique-t-il? Un certain laxisme, un peu de nonchalance des répondants dans des sous-tests assez courts pourrait expliquer leur performance inférieure; deux professeurs qui ont collaboré avec nous ont dit qu'il leur a été difficile de faire passer les sous-tests à six items dans un climat de sérieux à cause de la courte durée de l'épreuve; il s'est développé dans leurs classes respectives une situation de "sprint" compétitif. En somme, un climat de démotivation parce que la tâche n'a plus d'importance aux yeux des répondants. Cette situation particulière ne se retrouve dans aucune des autres recherches et, de ce fait, devient une explication encore plus plausible. Si telle était l'explication, il serait facile de corriger pour éviter cela et ainsi procéder à la collecte de données par la technique ECIS sans sous-estimer la moyenne.

Dans l'ensemble, les recherches sur l'effet de contexte font état d'une moindre influence de ce dernier sur l'estimation de la variance. Nos trois plans ont donné des estimateurs ECIS réel et ECIS simulé de la

variance toujours comparables. L'estimation ECIS de la variance apparaît donc assez imperméable à l'application de la technique ECIS à la collecte des données. Quant aux moments d'ordres 3 et 4, les recherches actuelles n'en font point mention, d'où l'impossibilité de confronter nos résultats à ceux d'autres chercheurs.

LE DEUXIEME PLAN EXPERIMENTAL

L'appendice C contient les listes d'indices de facilité de tous les items pour chaque échantillon, ainsi que les moyennes du premier et du dernier quarts. Pour chacun des plans ECIS réel ainsi que pour l'échantillon des 171 répondants aux 72 items, nous avons commencé par vérifier la similitude entre le premier et le dernier quarts du test intégral en effectuant une analyse de la variance des indices de facilité entre ces parties. Les tableaux 13 à 16 présentent le sommaire des analyses de la variance pour chacun des plans ECIS réel et pour l'échantillon des 171 répondants au test intégral. Tel que prévu par construction, les moyennes des indices de facilité des items des premier et dernier quarts du test intégral se sont avérées équivalentes dans tous les cas.

TABLEAU 13

Analyse de la variance des indices de facilité des items du premier et du dernier quarts du test intégral au plan ECIS (12,6) réel

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,0075	1	0,0075	0,1799
ERREUR	1,4195	34	0,0417	
TOTAL	1,4270	35		

TABLEAU 14

Analyse de la variance des indices de facilité des items du premier et du dernier quarts du test intégral au plan ECIS (4,18) réel

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,0038	1	0,0038	0,1043
ERREUR	1,2532	34	0,0369	
TOTAL	1,2570	35		

TABLEAU 15

Analyse de la variance des indices de facilité des items du premier et du dernier quarts du test intégral au plan ECIS (2,36) réel

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,0018	1	0,0018	0,0520
ERREUR	1,2078	34	0,0355	
TOTAL	1,2096	35		

TABLEAU 16

Analyse de la variance des indices de facilité des items du premier et du dernier quarts du test intégral pour les répondants aux 72 items

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,0017	1	0,0017	0,0573
ERREUR	0,9890	34	0,0291	
TOTAL	0,9907	35		

Nous avons comparé les indices de facilité calculés à partir des sous-tests de chaque plan ECIS à ceux calculés sur tout l'échantillon de 171 sujets qui a servi à la simulation des plans ECIS; vu le nombre de répondants, nous croyons que ces indices sont les plus sûrs que nous ayons sur chaque item pour la population en cause. Puisque ici les mêmes items, comme sujets, ont été deux fois répondus, nous sommes dans un cas de mesures dites répétées, d'où la nécessité d'analyser la variance en tenant compte de la corrélation entre les sources de variation, ce qui donne lieu à un type particulier d'analyse de la variance.

QUESTION 1

Etant donné un contexte de normalisation d'un test de 72 items par un plan ECIS (12,6), pour les 18 premiers items du test intégral, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

Les moyennes des indices de facilité sont respectivement de 0,46 et 0,55 aux sous-tests et au test. L'analyse de la variance des indices de

facilité révèle une différence significative à $p \leq 0,002$ (tableau 17) qui appelle comme conséquence une réponse affirmative à la question posée. On peut donc émettre l'hypothèse que la performance des répondants aux 18 premiers items d'un test qui en compte 72 ne sera pas la même suivant qu'ils répondent à ces items dans un contexte de test à 72 items ou dans un contexte de test à six items; qui plus est, leur performance serait réduite dans le contexte de tests courts.

TABLEAU 17

Analyse de la variance des indices de facilité des items 1 à 18 répondus dans les sous-tests de six items et dans le test intégral

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,070	1	0,070	19,6290 *
ITEMS	1,119	17	0,066	
(ERREUR)	0,061	17	0,004	
TOTAL	1,250	35		

* Rapport significatif à $p \leq 0,002$

QUESTION 2

Etant donné un contexte de normalisation d'un test de 72 items par un plan ECIS (12,6), pour les 18 derniers items de ce test, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

A $p \leq 0,05$, les moyennes 0,53 (répondants au test) et 0,49 (répondants aux sous-tests) ne peuvent être considérées significativement différentes (voir tableau 18). Par conséquent, on peut retenir l'hypothèse que le rendement des répondants à ces items est peu influencé par les mo-

difications au contexte que créent les sous-tests de six items.

TABLEAU 18

Analyse de la variance des indices de facilité des items 55 à 72 répondus dans les sous-tests de 18 items et dans le test intégral

SOURCE	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,019	1	0,019	2,4175
ITEMS	1,096	17	0,064	
(ERREUR)	0,132	17	0,008	
TOTAL	1,247	35		

QUESTION 3

Etant donné un contexte de normalisation d'un test de 72 items par un plan ECIS (4,18), pour les 18 premiers items du test intégral, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

Les moyennes des indices de facilité des deux groupes sont 0,55 (répondants au test) et 0,51 (répondants aux sous-tests); voir tableau 19. L'analyse de la variance des indices de facilité compilés pour les deux groupes amène à la conclusion que ces deux séries d'indices sont comparables et que, par conséquent, le sont aussi les rendements moyens des deux groupes expérimentaux.

TABLEAU 19

Analyse de la variance des indices de facilité des items 1 à 18 répondus dans les sous-tests de 18 items et dans le test intégral

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,014	1	0,014	2,4388
ITEMS	0,973	17	0,057	
(ERREUR)	0,098	17	0,006	
TOTAL	1,086	35		

QUESTION 4

Etant donné un contexte de normalisation d'un test de 72 items par un plan ECIS (4,18), pour les 18 derniers items de ce test, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

Les moyennes des indices sont de 0,53 (répondants au test) et 0,49 (répondants aux sous-tests) et l'analyse de la variance (tableau 20) conclut à l'hypothèse nulle. En conséquence, il ne semble pas, encore ici, que le rendement des répondants ait été perturbé par la situation des sous-tests ECIS (4,18).

QUESTION 5

Etant donné un contexte de normalisation d'un test de 72 items par un plan ECIS (2,36), pour les 18 premiers items du test intégral, le rendement des répondants aux sous-

TABLEAU 20

Analyse de la variance des indices de facilité des items 55 à 72 répondus dans les sous-tests de 18 items et dans le test intégral

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,014	1	0,014	2,4388
ITEMS	0,973	17	0,057	
(ERREUR)	0,098	17	0,006	
TOTAL	1,086	35		

tests sera-t-il significativement différent de celui des répondants au test intégral?

Les indices issus du plan ECIS (2,36) ont une moyenne de 0,51, tandis que ceux issus du test ont une moyenne de 0,55. L'analyse de la variance (tableau 21) ne permet pas de rejeter l'hypothèse nulle. Nous en déduisons que les deux groupes de répondants ont eu une performance semblable aux 18 premiers items du test.

QUESTION 6

Etant donné un contexte de normalisation d'un test de 72 items par un plan ECIS (2,36), pour les 18 derniers items de ce test, le rendement des répondants aux sous-tests sera-t-il significativement différent de celui des répondants au test intégral?

Les deux moyennes des indices sont égales à 0,53, évidemment comparables selon l'analyse de la variance (tableau 22). Nous conservons donc

TABLEAU 21

Analyse de la variance des indices de facilité des items 1 à 18 répondus dans les sous-tests de 36 items et dans le test intégral

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,009	1	0,009	1,8702
ITEMS	1,050	17	0,062	
(ERREUR)	0,084	17	0,005	
TOTAL	1,143	35		

l'hypothèse que le rendement des répondants aux items 55 à 72 n'est pas sensiblement perturbé par les modifications ECIS au contexte.

TABLEAU 22

Analyse de la variance des indices de facilité des items 55 à 72 répondus dans les sous-tests de 36 items et dans le test intégral

Source	Somme des carrés	Degrés de liberté	Carrés moyens	F
GROUPES	0,000	1	0,000	0,0220
ITEMS	0,949	17	0,056	
(ERREUR)	0,114	17	0,007	
TOTAL	1,063	35		

DISCUSSION

Le tableau 23 fait état de la faible variation de la moyenne des indices de facilité des 18 premiers items et des 18 derniers, pour chacun des plans ECIS réel et pour les répondants au test intégral. Même si l'analyse de la variance nous autorise à retenir l'hypothèse que, au premier quart du test, les répondants ont moins bien réussi en situation de sous-tests à six items qu'en situation de test à 72 items, l'ensemble des résultats invite à la prudence pour ce même plan ECIS, car la performance des répondants aux derniers items ne subit pas de fluctuation significative. Bien que l'analyse des recherches antérieures ne nous ait pas conduit à formuler des hypothèses fermes en regard des derniers items du test, notre expectative allait dans le sens d'une différence entre les moyennes, plus grande pour les derniers items que pour les premiers. Compte tenu de cette argumentation, il nous apparaît plus prudent de supposer que cet écart significatif provient d'erreurs expérimentales.

TABLEAU 23

Les moyennes des indices de facilité des items 1 à 18 et 55 à 72 pour les échantillons de sujets des plans ECIS réel et pour l'échantillon de sujets du test à 72 items

Items	Test à 72 items	Plan ECIS réel (2,36)	(4,18)	(12,6)
1 à 18	0,55	0,51	0,51	0,46
55 à 72	0,53	0,53	0,49	0,49

Il ne faut cependant pas s'arrêter là et observer que tous les plans ECIS réel donnent des indices plus faibles que ceux tirés du test à 72 items. Tout particulièrement, les plans (4,18) et (12,6) donnent tous une valeur inférieure à 0,55 comme moyenne des indices, d'où, peut-être,

une légère tendance à une performance moindre des répondants dans ces contextes. Mais là encore, étant donné les travaux antérieurs, il semble qu'il faille attribuer cette constatation mineure à des erreurs expérimentales.

Sous réserve de ses limites, cette partie de notre étude pour les trois plans ECIS expérimentés révèle que, autant pour les premiers items d'un test que pour les derniers, on ne remarque, à travers les indices de facilité, aucune modification sensible de la performance des répondants autre que celle due, selon toute vraisemblance, à une erreur expérimentale.

La question de base à laquelle ce plan voulait apporter des réponses se lit: *les modifications ECIS au contexte influencent-elles différemment le rendement des répondants aux items selon que ceux-ci se trouvent au début ou à la fin du test intégral?*

CONCLUSION

Après avoir résumé la recherche, nous présentons des considérations diverses sur le travail fait, notamment des critiques, une appréciation de la méthodologie et quelques suggestions de travaux futurs. Nous fermons avec quelques recommandations au sujet de la technique d'échantillonnage combiné des items et des sujets (ECIS).

RESUME

Normaliser un test pour une population consiste à déterminer les paramètres de la performance collective de tous les membres de cette population à ce test. En recueillant des données fragmentaires par administration de sous-tests issus du test intégral à autant de sous-échantillons de la population, la technique d'échantillonnage combiné des items et des sujets (ECIS) estime les paramètres de la population parente. A ce jour, les chercheurs, par simulation de la technique ECIS, ont prouvé qu'elle produit des estimateurs aussi fiables que ceux des autres techniques. Cependant, les occasions où elle a vraiment servi à la collecte des données sont plus rares et l'hypothèse d'un effet dû au contexte ECIS d'application sur le rendement des répondants inquiète toujours les utilisateurs.

En effet, lorsque les items du test intégral sont distribués dans les sous-tests, plusieurs caractéristiques du contexte d'application de l'épreuve subissent des changements, notamment la réduction de la durée de l'épreuve et le réarrangement des items. Jusqu'ici, les recherches sont peu concluantes.

En étudiant cet effet du contexte ECIS, nous voulons répondre, à la lumière de résultats expérimentaux, aux questions suivantes: 1°) *Quelle influence les modifications ECIS au contexte ont-elles sur les estimations de la moyenne, de la variance et des moments d'ordre 3 et 4?* 2°) *Les modifications ECIS au contexte influencent-elles différemment le rendement des répondants aux items selon que ces derniers se trouvent au début ou à la fin du test intégral?*

La première question de base donne lieu à un premier plan expérimental où la variable indépendante est la technique de collecte des données et la variable dépendante l'estimation des quatre premiers moments de la distribution. Les estimations ainsi obtenues furent comparées aux estimations issues de plans ECIS simulé correspondants.

Dans la deuxième partie de notre recherche, nous tentons de vérifier l'hypothèse selon laquelle, lors de l'application de la technique ECIS à la collecte des données, les changements contextuels relatifs aux derniers items auraient plus d'influence sur la performance des répondants que ceux des premiers items. Nous avons simplifié l'expérience en ne considérant que les premier et dernier quarts du test, soit les items 1 à 18 et 55 à 72. Les indices de facilité des items servirent de variables dépendantes, postulant que leur variation est directement proportionnelle à celle de la performance des répondants.

Les deux plans expérimentaux furent intégrés dans une seule expérimentation. Tout d'abord, nous avons construit un test de 72 items à choix multiple extraits de deux versions d'un test d'intelligence déjà validé pour une population du même âge que la nôtre. Les indices de facilité des items étant connus, au moment de leur sélection nous avons rendu équivalents les premier et dernier quarts du test par appariement des items.

Les élèves de secondaire I de quatre commissions scolaires du Québec formèrent la population de référence. Le test fut normalisé avec les trois plans ECIS, soit (12,6), (4,18) et (2,36), sur trois échantillons représentatifs distincts. Concurremment, un autre échantillon représentatif passa le test intégral; c'est avec cet échantillon que furent simulés les plans ECIS.

En recueillant les données avec les plans ECIS (12,6) et (4,18), dans les deux cas les estimateurs ECIS réel et ECIS simulé de la moyenne s'écartaient significativement l'un de l'autre. Avec le plan ECIS (2,36), les deux estimateurs semblaient estimer la moyenne d'une même population. Pour estimer tant la variance que le moment d'ordre 3, quand elle est appliquée véritablement à la collecte des données, la technique ECIS semble fiable; du moins, nos résultats n'infirment pas cet avancé. Préférant attribuer à une erreur d'échantillonnage l'écart significatif constaté pour le plan (2,36), nous pensons que le moment d'ordre 4 peut aussi être estimé avec sûreté par l'utilisation d'ECIS.

Quant à la variation des indices de facilité, puisqu'une seule différence significative a été constatée dans le cas du plan ECIS (12,6) réel, il apparaît préférable de l'attribuer à des sources expérimentales d'erreur. On peut donc raisonnablement supposer que l'application de la technique ECIS n'a pas d'effet localisé sur la performance des répondants, dans des conditions analogues à celles de notre expérience.

Cette recherche tend à confirmer les études antérieures selon lesquelles l'estimation de la variance par la technique ECIS est suffisamment fiable. Nous retenons aussi l'hypothèse que l'on peut estimer sans biais significatif le moment d'ordre 3; certes l'absence d'études pour la corroborer nous invite à la prudence. Ceci soulève la question du plan ECIS idéal pour éviter de biaiser l'estimation de la moyenne; tenant compte des suggestions de Shoemaker (1973b, pages 21 à 23) pour un test de 72 items comme le nôtre, des plans ECIS (6,12) ou (8,9) sans répétition d'items dans un même test ou, encore mieux, des plans (8,18) ou (9,16) à items répétés

dans plusieurs tests auraient de bonnes chances d'éviter l'écueil de l'effet de contexte.

Par ailleurs, le réarrangement des items et la variation de la durée de l'épreuve ne seraient pas des facteurs influençant sensiblement la performance des répondants; si une variation de performance aux items se produisait en passant du test au sous-test, celle-ci se manifesterait à peu près uniformément sur l'ensemble des items du test.

CONSIDERATIONS SUR LA RECHERCHE

LA METHODOLOGIE

Sur le plan de la méthodologie, cette étude innove de trois façons: d'abord, par ses quatre opérations distinctes et simultanées de normalisation d'un test à une même population; ensuite, par la comparaison des estimateurs ECIS réel à leurs correspondants ECIS simulé; enfin, par la partition du test et l'équilibrage de ses premier et dernier quarts aux fins d'étude de la variation des indices de facilité. A l'origine, ces innovations nous sont apparues nécessaires pour atteindre les objectifs fixés et établir les contrôles requis; ce sont des choix qui s'inspiraient de l'état de la recherche et accordaient crédit aux résultats observés. Conscient des risques que comportaient ces choix, nous fûmes attentif, tout au cours de cette recherche, aux conséquences qui pouvaient en résulter. Nos *critiques*, nos *perspectives* et même nos *recommandations* au sujet de la technique sont imprégnées de cet esprit de prudence.

CRITIQUES

Dans notre deuxième plan expérimental, nous avons postulé que la variation de la performance des répondants peut se mesurer par la variation des indices de facilité des items. Nous avons raisonné selon une logique dont les termes peuvent être ainsi énoncés: 1°) le niveau de difficulté d'un item est fonction de l'ordre de complexité du problème (composante intrinsèque) qu'il pose au répondant, et de l'habileté potentielle de celui-ci à résoudre le problème (composante extrinsèque); 2°) la perfor-

mance du répondant audit item est fonction de son habileté potentielle à résoudre le problème que lui soumet l'item (composante intrinsèque), et des contingences contextuelles dans lesquelles il est placé pour résoudre le problème (composante extrinsèque); 3°) d'où, si deux échantillons représentatifs d'une même population répondent à un même item dans des conditions contextuelles différentes et que l'on constate un niveau moyen de performance distinct pour chacun, cette différence sera attribuable aux distinctions caractéristiques des deux contextes de réponse; et, 4°) si la réponse est unique, bonne ou fausse, l'indice de facilité de l'item, rapport du nombre de bonnes réponses sur le nombre total de répondants à l'item, devient l'indice de la performance moyenne des répondants à ce dernier et toute variation de performance à un item se traduira par une variation de son indice de facilité.

Nous croyons devoir émettre quelques réserves à ce raisonnement essentiellement logique. En utilisant une mesure indirecte de la performance, on perd inévitablement, par dégénérescence, certains renseignements donnés par une mesure directe comme le résultat individuel; par exemple, l'indice de facilité nous renseigne davantage sur le nombre de répondants. Nous n'avons pas évalué l'incidence de cette perte de précision sur la finesse d'analyse des tests statistiques de comparaison, mais il serait opportun de le faire avant de reprendre le modèle pour une autre étude.

Lorsqu'un expérimentateur veut travailler avec des populations de sujets de niveau scolaire, il n'a souvent guère la possibilité de constituer des échantillons représentatifs en tirant au sort les sujets. Nous ne faisons pas exception, car effectivement les sujets de nos échantillons appartiennent aux groupes-classes choisis; en quelque sorte, nous avons procédé à un échantillonnage "en grappe". On sait les inconvénients que cette méthode comporte; par exemple, il serait possible que l'un des plans ait été expérimenté par hasard uniquement dans des groupes-classes homogènes forts ou faibles. A certains égards, nous n'avions aucun contrôle direct là-dessus sauf le jugement des professeurs-collaborateurs. Les résultats n'éveillent aucun soupçon... Mais!

Notre approche méthodologique nous a amené à comparer des estimateurs

ECIS réel à leurs correspondants ECIS simulé. Jusqu'ici, il n'existe pas, à notre connaissance, de précédent. Il suffit que les deux estimateurs (ECIS réel et ECIS simulé) biaisent dans le même sens pour que l'on en conclue qu'ils sont fiables alors qu'ils pourraient bien ne pas l'être, l'un ou l'autre ou les deux. Quoique nous ne l'ayons pas incorporé à notre recherche, nous avons effectué une comparaison par test "t" et test "F" des estimateurs ECIS réel et simulé de la moyenne et de la variance avec les estimateurs tirés des résultats des 171 répondants au test intégral. Nous avons constaté les mêmes divergences et convergences que celles révélées par nos tests expérimentaux; nous croyons cette constatation réconfortante, mais il ne s'agit en fait que d'une vérification valable pour le cas actuel seulement. S'il y avait eu divergence, qu'aurions-nous dit?

PERSPECTIVES

Pour obvier à certaines faiblesses inhérentes à notre expérience, nous croyons qu'une étude s'appuyant sur une adaptation du plan expérimental Solomon à quatre groupes (Van Dalen, 1966) clarifierait certaines zones obscures de notre plan en contrôlant et en comparant des mesures répétées sur les mêmes échantillons et d'autres, exclusives. Le plan aurait le déroulement suivant: 1º) une fois le test et la population choisis, sélection de quatre échantillons représentatifs de sujets; 2º) création d'un plan ECIS approprié; 3º) première phase expérimentale, un échantillon A subit les sous-tests du plan ECIS, deux échantillons B et C subissent le test intégral; 4º) deuxième phase expérimentale, l'échantillon A subit le test, l'échantillon C et un échantillon D subissent les sous-tests; 5º) simulation du plan ECIS sur les échantillons A et B; 6º) comparaison des six estimateurs (A_1, B_1, C_1, A_2, C_2, D_2) de chaque moment entre eux. Ce plan aurait le mérite de contrôler plusieurs facteurs expérimentaux dont nous avons présumé les effets compensatoires.

Même si les dernières recherches sur l'effet du contexte ECIS remontent à quelques années, les suggestions de leurs auteurs à l'effet d'étudier la validité de toutes les statistiques sont encore d'actualité. Nous avons ignoré les coefficients de consistance interne (formules 20 et 21 de Kuder-Richardson), le coefficient de fiabilité de Spearman-Brown, les

indices d'asymétrie et d'aplatissement, la distribution de fréquences et, pourtant, ce sont toutes des statistiques intéressantes à établir au moment de l'étalonnage complet d'un test.

Les limites mentionnées au premier chapitre de cette thèse invitent à varier les sujets, les types de tests, les plans ECIS, les types d'items, dans d'autres études. Par exemple, la validation d'un questionnaire d'attitudes par la technique ECIS résisterait-elle aux changements de contexte? Rien n'a encore été fait dans cette direction.

RECOMMANDATIONS AU SUJET DE LA TECHNIQUE ECIS

Quelques retombées de cette étude peuvent s'exprimer sous forme de recommandations à l'utilisateur de la technique d'échantillonnage combiné des items et des sujets. En voici quelques-unes que nous croyons directement dérivées de notre étude:

1º) Si l'on veut éviter une estimation biaisée de la moyenne, il est préférable de choisir un plan ECIS tel que les sous-tests aient une taille qui conserve à la tâche un caractère motivant pour les répondants. Par exemple, pour un test de 72 items, on peut croire que des sous-tests de 18 ou 24 items seraient convenables.

2º) Au moment de construire le test, il n'y a pas lieu de se préoccuper de l'arrangement des items dans le test en regard de l'utilisation d'un échantillonnage combiné des items et des sujets.

3º) Cette étude renforce l'hypothèse que l'estimation de la variance et du moment d'ordre 3 par la technique d'échantillonnage combiné des items et des sujets est fiable quel que soit le plan choisi.

4º) En milieu scolaire québécois de niveau secondaire, on peut distribuer les sous-tests à des échantillons établis par échantillonnage "en grappe", où les grappes sont des groupes-classes.

APPENDICE A

LE TEST ET LES SOUS-TESTS

TEST D'HABILETE MENTALE

NE PAS OUVRIR, NE PAS TOURNER LES PAGES

LE PROFESSEUR VA T'INDIQUER COMMENT REMPLIR LA FEUILLE DE REPONSE

En répondant sérieusement aux questions de cet examen, tu nous aideras à connaître leur difficulté et à savoir leur utilité. Ce sont toutes des questions avec un choix de quatre ou cinq réponses; tu choisis celle que tu crois être la bonne réponse. Avec ton crayon à mine, vis-à-vis le numéro de la question, tu noircis le rectangle sous la lettre correspondant à celle de la réponse choisie.

POUR LA QUESTION 78:

Laquelle des cinq choses suivantes est molle?

(f) le verre, (g) la pierre, (h) la ouate,

(j) le fer, (k) la glace.

Evidemment, la réponse est la ouate (h).

Alors vis-à-vis 78, tu noircis le rectangle sous:

I M P O R T A N T

Répondre vis-à-vis le bon numéro. ATTENTION, les réponses aux questions impaires sont lettrées a, b, c, d et é et celles aux questions paires f, g, h, j e. k.

Répondre aux questions dans leur ordre; ne pas revenir en arrière.

Si tu changes ta réponse, efface la première.

Ton examen comprend 72 questions; il est important que tu répondes vis-à-vis le numéro correspondant à celui de la question (Ex.: Si ta première question est numérotée 49, indique ta réponse vis-à-vis 49).

Dès que tu as terminé, écris le temps pris à répondre aux questions, sur la feuille de réponse, à droite du mot LEMMON.

Je sais que tu es sérieux et que tu vas répondre avec application aux questions de cet examen. Je t'en remercie très sincèrement.

Roger Bédard

ATTENDRE QUE LE PROFESSEUR DISE DE COMMENCER

1. Un lit a toujours ...

   (a) un ciel (b) une longueur (c) un oreiller (d) des couvertures
   (e) des ressorts

---

2. Dans la série de chiffres suivante, si l'on met des 9 à la place des 8 et qu'on enlève les 7, quel chiffre viendra après le septième 9?

   7 9 5 9 8 5 6 9 8 7 9 8 7 8 6 9 9 9 5

   (f) 5 (g) 6 (h) 7 (j) 9 (k) 8

---

3. Si les mots suivants sont disposés de façon à former la meilleure phrase possible, par quelle lettre commencera le dernier mot?

   brindilles un ramasse oiseau des nid son pour

   (a) b (b) r (c) o (d) n (e) s

---

4. Si Carole est plus petite qu'Estelle et Estelle de la même grandeur que Marie, alors Marie est (?) que Carole ...

   (f) plus petite (g) de la même grandeur (h) impossible à dire
   (k) plus grande

---

5. Si les mots suivants sont placés de façon à former la meilleure phrase possible, par quelle lettre commencera le dernier mot?

   tombe nuages des la pluie gris

   (a) t (b) n (c) d (d) g (e) p

---

6. Le verre est à la fenêtre ce que le ciment est ...

   (f) au sable (g) au trottoir (h) à dur (j) à la porte
   (k) au transparent

---

7. L'oeil est à la vue ce que l'ouie est ...

(a) aux lunettes (b) à la voix (c) au son (d) à l'oreille
(e) aux écouteurs

---

8. Si Hélène est plus vieille que Claire et si Hélène est plus jeune qu'Alice, alors Alice est (?) que Claire?

(ó) plus jeune (g) plus vieille (h) du même âge (j) impossible à dire

---

9. Trois des quatre dessins de droite ont quelque chose de semblable. Lequel diffère des trois autres?

(a) (b) (c) (d)

---

10. Un panier est à des oeufs ce qu'une enveloppe est à ...

(ó) une poignée (g) des oranges (h) des timbres (j) une lettre
(k) un facteur

---

11. D'un objet qui n'est pas aussi bon qu'un autre, on dit qu'il est...

(a) impropre (b) moindre (c) simple (d) inférieur (e) superflu

---

12. Quatre des mots suivants ont quelque chose de commun. Lequel est différent des quatre autres?

(ó) égoiste (g) malhonnête (h) pauvre (j) avare (k) grossier

---

13. Un éducateur ne doit jamais (?) un élève pour ses fautes.

(a) punir (b) réprimander (c) excuser (d) ridiculiser (e) aider

---

14. Dans une certaine rangée d'arbres, un arbre est le sixième de chaque extrémité de la rangée. Combien d'arbres y a-t-il dans la rangée?

(f) 6 (g) 11 (h) 10 (j) 12 (k) 13

---

15. Lequel des cinq objets suivants ressemble le plus à une clé, une pièce de monnaie ou un boulon?

(a) un billet d'un dollar (b) une porte (c) une valise
(d) une cuillère (e) une assiette

---

16. Meilleur est à bon ce que (?) est à mauvais.

(f) moindre (g) moins (h) détestable (j) pire (k) terrible

---

17. Au moment où une horloge marque 10 heures et 9 minutes, on interchange la position des deux aiguilles. L'horloge indique alors ...

(a) 1 heure et 9 minutes (b) 11 heures moins 9 minutes
(c) 2 heures et 9 minutes (d) 2 heures moins 9 minutes

---

18. J'ai une grande boîte dans laquelle il y a 4 petites boîtes qui contiennent chacune 3 boîtes très petites. Combien ai-je de boîtes en tout?

(f) 7 (g) 12 (h) 13 (j) 16 (k) 17

---

19. L'écorce de l'arbre est à l'arbre ce que la pelure de la banane est...

(a) à une feuille d'arbre (b) à la peau de banane (c) à l'écorce terrestre (d) à la banane (e)au bananier

---

20. Lequel des cinq objets suivants ressemble le plus à une cheminée, à un toit et à une porte?

(f) une chaise (g) un lit (h) un poêle (j) une fenêtre (k) un pupitre

---

21. Qu'est-ce qui exprime le mieux ce qu'est une exagération?

(a) un mensonge (b) une réponse fausse (c) une amplification de la vérité (d) une diminution de la vérité (e) une fausse déclaration

---

22. Si l'on met en ordre ce qui suit, qu'est-ce qui serait au milieu?

(f) chopine (g) baril (h) tasse (j) pinte (k) gallon

---

23. Dans une langue étrangère, *boli deta kipo* signifie *très beau temps; boli cora* signifie *mauvais temps;* et *deta sedu* signifie *très chaud.* Quel mot signifie beau?

(a) boli (b) deta (c) cora (d) kipo (e) sedu

---

24. Trois des quatre dessins de droite ont quelque chose de semblable. Lequel est différent des trois autres?

(f) (g) (h) (j)

---

25. Un homme a parcouru 9 milles vers l'est, à partir de sa maison puis 4 milles vers l'ouest. Il était alors rendu (?) de sa maison?

(a) 5 milles à l'est (b) 5 milles à l'ouest (c) 13 milles à l'est
(d) 13 milles à l'ouest

---

26. Une chaise possède nécessairement ...

(f) des bascules (g) du rembourrage (h) des barreaux
(j) un siège (k) des bras

---

27. Un pot est à la crème ce qu'un bol est à ...

(a) à la bouteille (b) à la soucoupe (c) à la tasse
(d) au sucre (e) à l'assiette

---

28. Compter les fois que le chiffre 4 est immédiatement suivi du chiffre 2 dans la ligne suivante?

2 4 1 4 2 3 5 4 6 2 4 7 5 2 4 4 2 3 9 4 3 2 8 7 8 4 2 2 4 5 5 2 2 4 2

Combien de fois? (f) 1 fois (g) 2 fois (h) 3 fois (j) 4 fois (k) 5 fois

---

29. Un général est à l'armée ce qu'un amiral est ...

(a) au commandant (b) à la guerre (c) à la marine (d) à l'océan
(e) à la feuille d'érable rouge

---

30. La copie au carbone d'une lettre s'appelle ...

(f) un multiple (g) une contrefaçon (h) un duplicata
(j) une imitation (k) un carbonate

---

31. J'ai une grande boîte dans laquelle il y a 4 petites boîtes. Chaque petite boîte contient 4 très petites boîtes. Combien ai-je de boîtes en tout?

(*a*) 8 (*b*) 9 (*c*) 21 (*d*) 16 (*e*) 20

---

32. Il y a un dicton: "Epargnez le fouet et vous gâtez l'enfant". Cela signifie...

(*f*) ne punissez jamais l'enfant

(*g*) les enfants seront gâtés si on ne les punit jamais

(*h*) les punitions gâtent l'enfant

(*j*) un fouet n'est pas un cadeau à donner à un enfant.

---

33. Laquelle des cinq personnes suivantes ressemble le plus à un avocat, à un pasteur et à un médecin?

(*a*) un commis (*b*) un policier (*c*) un charpentier (*d*) un professeur

(*e*) un laitier

---

34. L'un des nombres de la série ci-dessous est incorrect. Quel devrait être le nombre correct?

1 2 4 8 16 32 48 128

(*f*) 96 (*g*) 6 (*h*) 64 (*j*) 12 (*k*) 24

---

35. La tante de la petite-fille de ta grand-mère est...

(*a*) ta soeur (*b*) ta tante (*c*) ta fille (*d*) ta nièce (*e*) toi-même

---

36. Laquelle de ces réponses exprime le mieux ce qu'est un rateau?

(*f*) une machine (*g*) un outil (*h*) un ustensile (*j*) une chose

(*k*) un meuble

---

37. Si l'on met en ordre ce qui suit, qu'est-ce qui serait au milieu?

(a) une cabane (b) un gratte-ciel (c) une maison (d) un château
(e) un chalet

---

38. Dans ce qui suit, qu'est-ce qui exprime le mieux ce qu'est une cuisine?

(f) une salle où l'on apprête les aliments
(g) un endroit où l'on garde les couteaux et les fourchettes
(h) une pièce de la maison
(j) une salle avec une table et des chaises
(k) une salle voisine de la salle à manger

---

39. L'oncle du petit-fils de ton grand-père est ...

(a) ton frère (b) ton fils (c) ton oncle (d) ton neveu (e) toi-même

---

40. Une déclaration conforme à ce qui s'est produit s'appelle ...

(f) une prévarication (g) une communication (h) une vérité
(j) un compte rendu (k) une accusation

---

41. Laquelle des choses suivantes ressemble le plus au thé, au lait et à la limonade?

(a) la moutarde (b) le vinaigre (c) le café (d) l'huile d'olive
(e) l'eau

---

42. Un pantalon a toujours ...

(f) une ceinture (g) des revers (h) des poches (j) un pli
(k) des coutures

---

43. Compter les fois que le chiffre 6 est immédiatement suivi du chiffre 8, dans la ligne suivante?

8 6 3 6 8 5 7 6 9 8 6 3 5 8 6 6 8 7 9 6 9 8 7 2 6 5 6 8 8 6 2 2 8 8 6 8

Combien de fois?

(a) 1 fois (b) 4 fois (c) 2 fois (d) 3 fois (e) 5 fois

---

44. Il y a un dicton: "Prévenir vaut mieux que guérir". Cela signifie que...

(f) la prévention est un bon remède (g) prévenir et guérir sont équivalents
(h) il est mieux de prévoir un malheur que de s'en sortir
(j) il est mieux de guérir que de prévenir

---

45. Qu'est-ce qui exprime le mieux ce qu'est un réfrigérateur:

(a) un meuble de cuisine (b) un endroit où l'on garde la nourriture
(c) un accessoire électrique pour la cuisine (d) une grande caisse blanche
(e) un cabinet où l'on conserve les aliments au froid

---

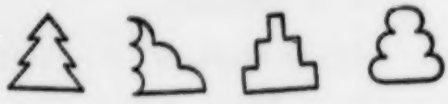
46. Lequel des cinq mots suivants ne va pas avec les autres:

(f) grand (g) honnête (h) fort (j) vigoureux (k) sain

---

47. Trois des quatre dessins de droite ont quelque chose de semblable. Lequel est différent des trois autres?

(a) (b) (c) (d)



---

48. Animal est à cheval ce que (?) est à carotte

(f) rouge (g) singe (h) repas (j) légume (k) lapin

---

49. Le froid est au réfrigérateur ce que la chaleur est ...

(*a*) à la température (*b*) au chaud (*c*) au poêle (*d*) à la cuisine
(*e*) à la glace

---

50. Laquelle de ces séries contient un nombre faux?

(*f*) 1-3-5-7-9 (*g*) 2-4-6-8-10 (*h*) 2-5-8-11-14 (*j*) 1-4-7-10-13
(*k*) 3-6-9-12-14

---

51. Le bois est à la boîte ce que le caoutchouc est ...

(*a*) au soulier (*b*) à l'élasticité (*c*) à l'eau (*d*) au pneu
(*e*) à la mousse

---

52. Dans une langue étrangère, *karo lupi tana* signifie *très petite fille*; *karo sali* signifie *grande fille* et *lupi neka* signifie *très bon*. Quel mot signifie *petit*?

(*f*) karo (*g*) lupi (*h*) sali (*j*) tana (*k*) neka

---

53. Le nord-ouest est au sud-est ce que le haut est...

(*a*) à plus haut (*b*) au dessous (*c*) à nord-est (*d*) au bas
(*e*) au nord

---

54. Une photographie mesure 3 pouces de largeur et 5 pouces de longueur. Si on l'agrandit à 12 pouces de largeur, quelle sera sa longueur?

(*f*) 8 pouces (*g*) 20 pouces (*h*) 14 pouces (*j*) 15 pouces
(*k*) 60 pouces

---

55. Dans ce qui suit, qu'est-ce qui ressemble le plus à une orange, une banane et une prune?

(a) un arbre (b) une pêche (c) une pelure (d) une noix
(e) une nourriture

---

56. Un garçon possède trois chiens. Ils ont pour noms Capi, Prince et Fido. Capi est plus gros que Prince et Prince plus gros que Fido. Alors, Capi est (?) que Fido.

(f) plus petit (g) plus gros (h) de la même grosseur
(j) impossible à dire

---

57. Mauvais est à pire ce que bon est à ...

(a) médiocre (b) mieux (c) meilleur (d) passable
(e) excellent

---

58. Laquelle de ces réponses exprime le mieux ce qu'est une théière?

(f) un outil (g) une arme (h) un ustensile (j) une fleur
(k) une machine

---

59. Si Thomas est plus brillant que Daniel et que Daniel est aussi brillant qu'Henri, alors Henri est (?) que Thomas.

(a) plus brillant (b) moins brillant (c) aussi brillant
(d) impossible à dire

---

60. L'estomac est à la nourriture ce que le coeur est ...

(f) à l'homme (g) aux veines (h) au sang (j) aux artères
(k) au battement

---

61. Dans une rangée d'arbres, un arbre est le cinquième de chaque extrémité de la rangée. Combien d'arbres y a-t-il dans la rangée?

(*a*) 5 (*b*) 6 (*c*) 10 (*d*) 9 (*e*) 11

---

62. L'un des chiffres de cette série est incorrect. Quel est le chiffre correct?

8 1 8 2 8 3 8 4 8 5 8 6 8 7 8 9

(*f*) 9 (*g*) 7 (*h*) 6 (*j*) 8 (*k*) 5

---

63. Le sol est à l'automobile ce que l'eau est ...

(*a*) au train (*b*) à la gazoline (*c*) au moteur (*d*) au navire

(*e*) au fleuve

---

64. Le maire d'une ville est à la ville ce que le premier ministre est ...

(*f*) au peuple (*g*) au gouverneur général (*h*) à la province

(*j*) au parlement (*k*) au cabinet

---

65. Si les mots suivants sont disposés de façon à former la meilleure phrase possible, quelle serait la première lettre du dernier mot?

plâtre faits quelquefois de sont les murs

(*a*) s (*b*) f (*c*) p (*d*) m (*e*) q

---

66. Laquelle des raisons suivantes est la plus importante pour acheter à tempérament?

(*f*) pour éviter de manipuler une grosse somme d'argent

(*g*) afin de pouvoir utiliser l'objet sans attendre d'avoir assez d'argent pour le payer entièrement

(*h*) c'est meilleur marché

(*j*) c'est plus commode ainsi et cela fait épargner de l'argent

---

67. Si l'on met en ordre ce qui suit, qu'est-ce qui sera au milieu?

(a) doigt (b) paume (c) poignet (d) épaule (e) coude

---

68. L'un des nombres de la série suivante est incorrect. Quel devrait être le nombre correct?

1 8 2 7 3 8 4 7 5 8 6 7 7 8 8 8

(f) 9 (g) 8 (h) 7 (j) 6 (k) 5

---

69. D'un objet qui n'est pas aussi bon qu'un autre, on dit qu'il est...

(a) impropre (b) moindre (c) simple (d) inférieur (e) superflu

---

70. Si les mots suivants sont disposés de façon à former la meilleure phrase possible, par quelle lettre commencerait le dernier mot?

plâtre faits souvent de sont murs les

(f) l (g) f (h) p (j) m (k) d

---

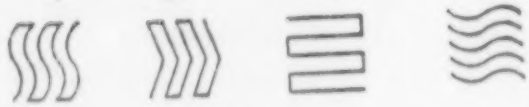
71. L'auteur est au livre ce que l'inventeur est...

(a) à la machine (b) au livret (c) au découvreur (d) à l'écrivain (e) au magazine

---

72. Trois des quatre dessins de droite ont quelque chose de semblable. Lequel est différent des trois autres?

(f) (g) (h) (j)



APPENDICE B

LES INSTRUCTIONS AU PROFESSEUR-COLLABORATEUR

AU PROFESSEUR-COLLABORATEUR

OBJET: Mode d'administration de l'examen

Madame,
Mademoiselle,
Monsieur,

Je veux à prime abord vous dire ma gratitude pour votre collaboration. L'entreprise que constitue une thèse de doctorat comporte de nombreuses embûches dont la difficulté d'obtenir une population auprès de laquelle expérimenter. C'est pourquoi, votre contribution m'est très chère et je vous en remercie.

Ma recherche porte sur la fatigue ressentie au cours d'examens et son influence sur le rendement de l'élève. Comme le sujet peut vous intéresser, je serai heureux de vous faire parvenir une copie du sommaire du rapport pourvu que, sur l'enveloppe, vous m'indiquiez votre nom et votre adresse. Evidemment, je vous invite à huit ou neuf mois de patience avant que j'en arrive là... mais je tiendrai ma promesse le moment venu.

En premier lieu, j'aimerais que vous adhériez strictement aux directives afin de maintenir les conditions expérimentales aussi identiques que possible d'une classe à l'autre; vous aiderez ainsi à réduire l'effet des fluctuations d'échantillonnage, condition indispensable à la validité des résultats.

LES DIRECTIVES

1- La veille du jour d'administration du questionnaire, inviter les élèves à se munir d'un crayon à mine (type HB) et d'une gomme à effacer.

N.B.: Je devrai retranscrire au crayon toutes les feuilles de réponse écrites à la plume... moins il y en aura, mieux ce sera. Merci!

2- Etablir un climat calme au début de la période et distribuer les questionnaires et les feuilles de réponse.

3- Indiquer aux élèves l'endroit où écrire leur nom en lettres moulées, noircir la lettre correspondant à leur sexe et les chiffres indiquant leur âge et la date où le questionnaire est administré.

N.B.: Vous avez un exemple annexé à la présente.

4- Donner aux élèves le code de l'école et leur faire noircir les chiffres appropriés dans la section prévue.

5- Dire aux élèves qu'ils n'ont pas à fournir d'autres renseignements.

N.B.: Les feuilles de réponse peuvent servir à d'autres questionnaires, c'est pourquoi d'autres renseignements peuvent y être consignés.

6- Lire et expliquer aux élèves les directives contenues sur la première feuille de leur questionnaire. Vous assurer que tous comprennent bien les directives.

7- Avertir les élèves qu'ils devront noter le temps pris pour répondre aux questions.

Suggestion: 1°) partir l'administration à la minute précise (Ex.: 14h 57sec 0min),

2°) inscrire le nombre de minutes écoulées à chacune,

3°) quand l'élève a terminé, il lève les yeux vers l'horloge et ne prend lecture que de la position de la trotteuse pour obtenir les secondes écoulées depuis la dernière minute inscrite au tableau.

8- Donner le signal du départ et indiquer le temps écoulé.

9- Recueillir les feuilles de réponse et les questionnaires.

10- Si, pour des raisons évidentes, vous jugez qu'un élève ne s'est pas conformé aux directives, faire une croix sur sa feuille et elle sera automatiquement rejetée.

11- Indiquer au dos de l'enveloppe (coin supérieur droit) l'heure à laquelle le questionnaire a été administré.

12- Retourner toutes les feuilles de réponse et tous les questionnaires dans l'enveloppe pré-adressée et pré-affranchie.

N.B.: Les feuilles de réponse et les questions utilisées sont assorties de copyright et ne peuvent donc être laissées en circulation libre. Merci de votre compréhension.

Remarquez que, selon le questionnaire qui est assigné à votre classe contient 6, 18, 36 ou 72 questions, le temps d'administration (sans compter les explications préliminaires) sera d'environ 3, 10, 15 ou 30 minutes.

En terminant, je ne peux que vous souhaiter la même collaboration si un jour vous faites une thèse et vous dire GRAND MERCI.

Roger Bédard

UNIVERSITÉ LAVAL
SCIENCE
DE L'ÉDUCATION

FEUILLE DE RÉPONSE OTIS-LENNON

2 min 15 sec
Le temps pris

UTILISER UN CRAYON À MINE — TYPE H-B

En lettres moulées

Répondre vis-à-vis la phrase

DIRECTIVES. Dans les cases ci-bas, inscrire en lettres moulées ou en chiffres tous les renseignements demandés, puis noircir les crochets qui correspondent à la lettre ou chiffre inscrit. N'inscrire qu'une seule lettre ou chiffre par case.

INSCRIRE VOTRE NOM, LAISSER UN ESPACE ET INSCRIRE VOTRE PRÉNOM

ROGER BEDARD

VOTRE CODE PERMANENT

SEXE

DATE DE NAISSANCE

JOUR	MOIS	AN	ÂGE
2	1	6 3	1 3

106

DATE DU TEST

JOUR	MOIS	AN
1 2	0 3	7 8

CODE DE L'ÉCOLE

E

INSCRIRE L'OCCUPATION DE VOTRE PÈRE OU VOTRE TUTEUR

INDIQUER LE NIVEAU DE SCOLARITÉ DE VOTRE PÈRE OU VOTRE TUTEUR

- MAÎTRISE OU DOCTORAT
- LICENCE OU BAC COMPLÈTE
- LICENCE OU BAC NON COMPLÈTE
- SECONDAIRE COMPLÈTE
- SECONDAIRE NON COMPLÈTE
- PRIMAIRE COMPLÈTE
- PRIMAIRE NON COMPLÈTE

USAGE OFFICIEL SEULEMENT

APPENDICE C

TABLEAU DES INDICES DE FACILITE
DES 72 ITEMS POUR CHAQUE ECHANTILLON

TABLEAU

Les indices de facilité des 72 items pour chaque échantillon

Item	Au test	(2,36)		(4,18)		(12,6)	
		Réel	Simulé	Réel	Simulé	Réel	Simulé
01	,285	,324	,162	,300	,290	,344	,313
02	,483	,378	,486	,200	,516	,344	,656
03	,738	,784	,784	,833	,806	,781	,875
04	,674	,757	,703	,533	,742	,594	,750
05	,698	,757	,784	,767	,645	,625	,688
06	,663	,595	,676	,667	,677	,594	,813
07	,552	,541	,514	,667	,548	,594	,688
08	,657	,568	,649	,733	,645	,531	,719
09	,442	,459	,351	,433	,516	,344	,563
10	,721	,676	,757	,500	,645	,531	,844
11	,651	,757	,595	,633	,806	,469	,625
12	,814	,649	,811	,633	,839	,813	,844
13	,442	,378	,432	,667	,548	,394	,424
14	,174	,135	,243	,200	,258	,061	,182
15	,576	,297	,541	,300	,645	,485	,515
16	,494	,432	,595	,433	,452	,242	,636
17	,424	,324	,378	,367	,484	,273	,576
18	,331	,432	,324	,367	,323	,212	,364
$\frac{\sum_{i=1}^{18} x_i}{18}$	0,55	0,51	0,54	0,51	0,58	0,46	0,62

Item	Au test	(2,36)		(4,18)		(12,6)	
		Réel	Simulé	Réel	Simulé	Réel	Simulé
19	,785	,757	,784	,697	,636	,733	,800
20	,227	,216	,243	,152	,455	,267	,133
21	,657	,568	,730	,485	,485	,600	,667
22	,494	,459	,459	,545	,667	,267	,467
23	,442	,324	,432	,394	,364	,200	,367
24	,547	,378	,459	,455	,848	,600	,533
25	,436	,459	,459	,424	,333	,387	,419
26	,849	,892	,892	,818	,758	,806	,742
27	,372	,378	,378	,273	,667	,226	,419
28	,756	,703	,757	,758	,182	,645	,806
29	,651	,676	,622	,545	,333	,516	,677
30	,180	,162	,108	,121	,636	,097	,194
31	,355	,459	,324	,242	,364	,129	,355
32	,628	,622	,676	,636	,545	,613	,484
33	,413	,297	,378	,394	,333	,290	,323
34	,512	,432	,568	,606	,788	,419	,484
35	,378	,351	,351	,333	,636	,452	,290
36	,837	,703	,839	,727	,545	,839	,677

Item	Au test	(2,36)		(4,18)		(12,6)	
		Réel	Simulé	Réel	Simulé	Réel	Simulé
37	,680	,679	,679	,571	,771	,613	,613
38	,610	,714	,571	,657	,686	,516	,606
39	,465	,429	,393	,457	,429	,452	,548
40	,273	,393	,286	,171	,314	,194	,258
41	,419	,464	,393	,286	,486	,355	,290
42	,849	,893	,857	,857	,914	,774	,903
43	,715	,607	,607	,600	,686	,724	,655
44	,581	,679	,500	,543	,600	,586	,552
45	,773	,821	,821	,857	,829	,862	,793
46	,494	,536	,429	,486	,314	,448	,448
47	,529	,464	,536	,229	,543	,483	,586
48	,798	,500	,679	,514	,771	,655	,759
49	,674	,786	,643	,714	,743	,556	,630
50	,663	,821	,679	,771	,771	,704	,630
51	,587	,571	,607	,686	,714	,593	,593
52	,419	,393	,321	,371	,371	,222	,370
53	,471	,607	,464	,400	,486	,370	,593
54	,378	,357	,357	,286	,343	,296	,556

Item	Au test	(2,36)		(4,18)		(12,6)	
		Réel	Simulé	Réel	Simulé	Réel	Simulé
55	,610	,429	,679	,444	,528	,533	,467
56	,593	,786	,714	,722	,583	,700	,500
57	,651	,464	,750	,417	,722	,533	,533
58	,651	,607	,714	,583	,583	,633	,667
59	,738	,679	,821	,778	,694	,600	,767
60	,279	,250	,286	,194	,278	,133	,300
61	,157	,321	,071	,056	,139	,029	,118
62	,506	,429	,321	,528	,389	,382	,500
63	,645	,679	,429	,611	,611	,765	,706
64	,599	,643	,429	,528	,583	,647	,676
65	,651	,571	,643	,500	,694	,412	,765
66	,273	,214	,321	,306	,306	,353	,235
67	,576	,643	,536	,472	,667	,364	,424
68	,320	,214	,250	,361	,167	,182	,182
69	,517	,750	,536	,583	,472	,576	,212
70	,698	,679	,571	,778	,722	,697	,364
71	,645	,679	,643	,444	,694	,545	,424
72	,465	,464	,393	,556	,472	,667	,455
$\frac{\sum_{i=55}^{72} x_i}{18}$	0,53	0,53	0,51	0,49	0,52	0,49	0,46

BIBLIOGRAPHIE

BARCIKOWSKI, Robert S., (1970a), Optimum use of the item sampling technique in obtaining test norms, rapport inédit présenté au Congrès annuel de l'American Educational Research en mars.

BARCIKOWSKI, Robert S., (1970b), Optimum use of the item sampling technique in obtaining test norms, Buffalo, State University of New York at Buffalo, thèse (Ph. D. Education).

BARCIKOWSKI, Robert S., (1972), A monte carlo study of item sampling (versus traditional sampling) for norm construction, *Journal of Educational Measurement*, 9, 3, p. 209-214.

BARCIKOWSKI, Robert S., (1974), The effect of item discrimination on the standard errors of estimate associated with item-examinee sampling procedures, *Educational and Psychological Measurement*, 34, p. 231-237.

BARCIKOWSKI, Robert S., OLSEN, H., (1975), Test item arrangement and adaptation level, *The Journal of Psychology*, 90, p. 87-93.

BARCIKOWSKI, Robert S., TERRANOVA, C., (1966), Item sampling, rapport inédit présenté au Congrès annuel de l'Educational Research Association of New York State en novembre.

BARCIKOWSKI, Robert S. et UPP, Caroline, (1976), Multiple matrix sampling step by step, Athens, Ohio, Ohio University, School of Applied Behavioral Sciences and Educational Leadership, manuscrit inédit.

BHUSHAN, Vidya, (1974), Adaptation of an intelligence test from English to French, *Journal of Educational Measurement*, 11, 1, p. 43-48.

BRANDENBURH, Dale C., FORSYTH, Robert A., (1973a), Approximating standardized achievement test norms with a theoretical model, rapport inédit présenté au Congrès annuel du National Council on Measurement in Education en février.

BRANDENBURG, Dale C., FORSYTH, Robert A., (1973b), The use of multiple matrix sampling to approximate norm distributions : An empirical comparison of two models, rapport inédit présenté au Congrès annuel du National Council on Measurement in Education en février.

BRENNER, Marshall H., (1964), Test difficulty, reliability, and discrimination as functions of item difficulty order, *Journal of Applied Psychology*, 48, 2, p. 98-100.

BURSACK, B.A., (1969), Utilizing item sampling techniques to scale affective reactions to mathematics, Ohio, Ohio State University, thèse (Ph. D.).

BURTON, N., REMER, R., (1972), Item sampling in questionnaires : rate of return and context effects, Colorado, University of Colorado, rapport no 57 du Laboratory of Educational Research, février.

CAHEN, L.S., ROMBERG, T.A., ZWIRNER, W., (1970a), The estimation of mean achievement scores for schools by the item-sampling technique, *Educational and Psychological Measurement*, 30, 1, p. 41-60.

CAHEN, L.S., ROMBERG, T.A., ZWIRNER, W., (1970b), A comparison of school mean achievement scores with two estimates of the same scores obtained by the item-sampling technique, Princeton, New Jersey, Educational Testing Service, *Research Bulletin RB-70-64*, novembre.

COOK, D.L., STUFFLEBEAM, D.L., (1967), Estimating test norms from variable size item and examinee samples, *Journal of Educational Measurement*, 4, p. 27-33.

CREHAN, Kevin D., KOEHLER, Roger A., SLAKTER, Malcom J., (1974), Longitudinal studies of test-wiseness, *Journal of Educational Measurement*, 11, 2, p. 209-212.

DIAMOND, J.J., EVANS, W.J., (1972), An investigation of the cognitive correlates of test-wiseness, *Journal of Educational Measurement*, 9, p. 145-150.

FELDT, L.S., FORSYTH, R.A., (1974), An examination of the context effect in item sampling, *Journal of Educational Measurement*, 11, p. 73-82.

FLAMENT, Claude, (1963), Fréquence et certitude des réponses dans une série prédictive, *l'Année Psychologique*, 63, p. 29-41.

FORESTER, Donald Lee, MICHAEL, William B., (1976), A comparison of the validities of a traditional item-allocation method and of an item-sampling procedure in the construction of in-class examinations in intermediate algebra, *Educational and Psychological Measurement*, 36, p. 933-938.

GIBB, B.G., (1964), Test-wiseness as secondary cue responses, Stanford University, thèse (Ph. D.).

GUILFORD, J.P., (1965), *Fundamental statistics in psychology and education*, quatrième édition, Toronto, McGraw-Hill.

HAMBLETON, R.K., ROVINELLI, R., GORTH, W.P., (1971), Efficiency of various item-examinee sampling designs for estimating test parameters, rapport inédit présenté au Congrès annuel de l'American Psychological Association en septembre.

HAMBLETON, R.K., TRAUB, Ross E., (1974), The effects of item order on test performance and stress, *The Journal of Experimental Education*, 43, 1, p. 40-46.

HILL, Richard K., (1975), Minimizing context effect when using multiple matrix sampling, rapport inédit présenté au Congrès annuel du National Council on Measurement in Education en avril.

HOOKE, R., (1953), Sampling from a matrix, with applications to the theory of testing, Princeton University, Statistical Research Group, *Memorandum Report* 53.

HUCK, Schuyler W., BOWERS, Norman D., (1972), Item difficulty level and sequence effects in multiple-choice achievement tests, *Journal of Educational Measurement*, 9, 2, p. 105-111.

IMPARA, J.C., (1972), An experimental comparison of matrix sampling and examinee sampling for estimating test norms for different target groups on different types of tests, The Florida State University, thèse (Ph. D.).

JACOBS, S.S., WILDEMANN, C., (1969), Matrix sampling with classroom tests : an empirical investigation, rapport inédit présenté au Congrès ERANYS en novembre.

JOHNSON, M.C., LORD, F.M., (1958), An empirical study of the stability of a group mean in relation to the distribution of test items among pupils, *Educational and Psychological Measurement*, 18, p. 325-329.

KLEINKE, David J., (1972), A linear-prediction approach to developing test norms based on matrix-sampling, *Educational and Psychological Measurement*, 1, p. 75-84.

KLEINKE, D.J.,(1973), Further studies of linear-prediction following matrix sampling, rapport inédit présenté au Congrès annuel conjoint A.E.R.A. - N.C.M.E. en février.

KLOSNER, Naomi Certner, GELLMAN, Estelle Klittnick, (1973), The effect of item arrangement on classroom test performance : implications for content validity, *Educational and Psychological Measurement*, 33, p. 413-418.

KNAPP, Thomas R., (1968), An application of balanced incomplete block design to the estimation of test norms, *Educational and Psychological Measurement*, 28, p. 265-272.

KNAPP, Thomas R., (1973), Item sampling, University of Rochester, texte inédit.

LOADMAN, W.E., (1972), an inquiry concerning the use of item sampling as a method to reduce testing time, Indiana University, Educational Research and Evaluation Laboratory, rapport no 5.

LORD, Frederic M., (1955), Sampling fluctuations resulting from the sampling of test items, *Psychometrika*, 20, 1, p. 1-22.

LORD, Frederic M., (1960), Use of true-score theory to predict moments of univariate and bivariate observed-score distributions, *Psychometrika*, 25, p. 325-342.

LORD, Frederic M., (1962), Estimating norms by item sampling, *Educational and Psychological Measurement*, 22, p. 259-267.

LYNCH, Daniel O., SMITH, Billie Clifton, (1975), Item response changes : effects on test scores, *Measurement and Evaluation in Guidance*, 7, 4, p. 220-224.

MILLMAN, J., BISHOP, H., EBEL, R., (1965), An analysis of test-wiseness, *Educational and Psychological Measurement*, (1965), 25, p. 707-726.

MISANCHUK, E.R., (1978), Multiple matrix sampling : a new technique for needs identification, *Canadian Journal of University Continuing Education*, 1978, 5, 1, p. 14-18.

MOLLENKOPF, William G., (1950), An experimental study of the effects on item-analysis data of changing item placement and test time limit, *Psychometrika*, 15, 3, p. 291-315.

MONTAGUE, M. Ann, (1971), Use of matrix sampling procedures with selected examinee and item populations to assess achievement in mathematics, The University of Wisconsin, thèse (Ph. D.).

MOORE, J.C., SCHUTZ, R.E., BAKER, R.L., (1966), The application of a self-instructional technique to develop a test-taking strategy, *American Educational Research Journal*, 3, p. 13-17.

MOY, Mabel L.Y, BARCIKOWSKI, Robert S., (1973), Item-sampling : optimum number of people and items, rapport inédit présenté au Congrès annuel du National Council on Measurement in Education en février.

MYERBERG, W. James, (1975), The effect of item stratification in multiple matrix sampling, rapport inédit présenté au Congrès annuel de l'American Educational Research Association en avril.

NEWELL, J.A., (1971), An empirical investigation of matrix sampling involving multiple item samples in a two-factor analysis of variance design, North Texas State University, thèse (Ph. D.).

NOIZET, G., FABRE, J.-M., (1975), Etude docimologique des questionnaires à choix multiple (Q.C.M.) : perspectives de recherche, *Scientia Paedagogica Experimentalis*, XII, 1, p. 38-62.

NOVAK, Carl D., (1974), An empirical investigation of multiple matrix sampling in a elementary school setting, rapport inédit présenté au Congrès annuel de l'American Educational Research Association en avril.

OWENS, T.R., STUFFLEBEAM, D.L., (1969), An experimental comparison of item sampling and examinee sampling for estimating test norms, *Journal of Educational Measurement*, 6, p. 75-83.

PEATMAN, John G., (1963), *Introduction to applied statistics*, New York, Harper & Row.

PIPER, R.M., SHOEMAKER, D.M., (1973), Formative evaluation of a kindergarten music program based on behavioral objectives, *Journal of Research in Music Education*, 21, p. 145-153.

PLUMLEE, L.B., (1964), Estimating means and standard deviations from partial data-an empirical check on Lord's item sampling technique, *Educational and Psychological Measurement*, 24, p. 623-630.

POGGIO, J.P., GLASNAPP, D.R., (1973a), Content sampling as an evaluation and research technique, *Educational and Psychological Measurement*, 33, p. 903-908.

POGGIO, J.P., GLASNAPP, D.R., (1973b), Item-sampling as a classroom evaluation technique, rapport inédit présenté au Congrès annuel du National Council on Measurement in Education en février.

PUGH, R.C., (1971), Empirical evidence on the application of Lord's sampling technique to Likert items, *Journal of Experimental Education*, 39, 3, p. 54-56.

QUEBEC (Province), ministère de l'Education, Service d'information, (1968), *Vocabulaire de l'éducation au Québec*, Québec, 68 p.

RASCH, G., (1966), An item analysis which takes individual differences into account, *The British Journal of Mathematical and Statistical Psychology*, 19, 1, p. 49-57.

REILE, Patricia J., BRIGGS, Leslie J., (1966), Should students change their initial answers on objective-type tests? : more evidence regarding an old problem, dans C.I. Chose et H.G. Ludlow, directeurs, *Readings in Educational and Psychological Measurement*, Boston, Houghton Mifflin Company, p. 371-375.

ROTHROCK, Julia E., MICHAEL, William B., (1976), The comparative validities of traditional item-selection and item-sampling procedures in evaluation of student and class progress toward course objectives

in freshman psychology, *Educational and Psychological Measurement*, 36, p. 925-932.

SAX, G., CROMACK, T.R., (1966), The effects of various forms of item arrangements on test performance, *Journal of Educational Measurement*, 3, p. 309-311.

SCHEETZ, Jim, (1976), Assigning items to subtests in multiple matrix sampling : an empirical post-mortem investigation, University of Iowa, thèse (Ph. D.).

SHOEMAKER, David M., (1970a), Allocation of items and examinees in estimating a norm distribution by item-sampling, *Journal of Educational Measurement*, 7, p. 123-128.

SHOEMAKER, David M., (1970b), Item-examinee sampling procedures and associated standard errors in estimating test parameters, *Journal of Educational Measurement*, 7, p. 255-262.

SHOEMAKER, David M., (1970c), Test statistics as a fonction of item arrangement, *Journal of Experimental Education*, 39, p. 85-88.

SHOEMAKER, David M., (1971a), An application of item-examinee sampling to scaling attitudes, *Journal of Educational Measurement*, 8, p. 279-282.

SHOEMAKER, David M., (1971b), Further results of the standard errors of estimate associated with item-examinee sampling procedures, *Journal of Educational Measurement*, 8, 3, p. 215-220.

SHOEMAKER, David M., (1971c), Principles and procedures of multiple matrix sampling, Southwest Regional Laboratory for Educational Research and Development, rapport technique no 34.

SHOEMAKER, David M., (1972), Standard errors of estimate in item-examinee sampling as function of test reliability variation in item difficulty indices and degree of skewness in the normative distribution,

Educational and Psychological Measurement, 3, p. 705-714.

SHOEMAKER, David M., (1973a), A note on allocating items to subtests in multiple matrix sampling and approximating standard errors of estimate with the jacknife, rapport inédit présenté au Congrès annuel du National Council on Measurement in Education en février.

SHOEMAKER, David M., (1973b), *Principles and Procedures of Multiple Matrix Sampling*, Cambridge, Mass., Ballinger Publishing Company.

SIROTNIK, Kenneth A., (1970), An investigation of the context effect in matrix sampling, *Journal of Educational Measurement*, 7, p. 199-207.

SIROTNIK, Kenneth A., WELLINGTON, Roger J., (1974), Scrambling content in achievement testing : an application of multiple matrix sampling in experimental design, *Journal of Educational Measurement*, 11, p. 179-188.

SLAKTER, M.J., KOEHLER, R.A., HAMPTON, S.H., (1970), Learning test-wiseness by programmed texts, *Journal of Educational Measurement*, 7, p. 247-254.

VAN DALEN, Deobold B., (1966), *Understanding Educational Research an Introduction*, Toronto, McGraw-Hill Book Company.

WHITELY, Susan E., DAWIS, René V., (1976), The influence of test context on item difficulty, *Educational and Psychological Measurement*, 36, p. 329-337.

WRIGHTSMAN, Lawrence S. Jr., (1962), The effects of anxiety, achievement motivation, and task importance upon performance on an intelligence test, *Journal of Educational Psychology*, 53, 3, p. 150-156.